Disproportionality Measures of Concentration, Specialization, and Localization

Frank Bickenbach, Eckhardt Bode

frank.bickenbach@ifw-kiel.de, eckhardt.bode@ifw-kiel.de

This version: May 29, 2008

Final version published in
International Regional Science Review
Volume 31 Number 4 (October 2008), pp. 359–388

Abstract

This paper extends the methodological toolbox of measures of regional concentration of industries and industrial specialization of regions. It first defines disproportionality measures of concentration and specialization, and proposes a taxonomy of these measures. This taxonomy is based on three characteristic features of any disproportionality measure. It helps researchers define the measure that fits their research purpose and data best. The paper then generalizes this taxonomy to cover disproportionality measures of economic localization that evaluate specialization and concentration simultaneously, and spatial disproportionality measures that deal with the checkerboard problem and the modifiable areal unit problem.

Keywords: *disproportionality measures, specialization, concentration, localization, spatial disproportionality measures*

The authors would like to thank Christiane Krieger-Boden, Serge Rey and two anonymous referees for helpful comments and suggestions, and the EU Commission for financial support.

1. Introduction

The new economic geography has raised concerns that economic integration at the regional and international level may increase 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 more slowly in terms of income and employment.

Various studies have explored the evolutions of concentration and specialization in Europe or 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 mainly two reasons (Combes and Overman 2004). First, many of these studies lack a clear-cut research purpose and test hypothesis, and choose their statistical measures largely ad hoc. They neglect the fact that choosing between different measures actually implies choosing between different *definitions* of concentration or specialization rather than just choosing between different ways of measuring a single, uniform theoretical construct of concentration or specialization.[2] Any inferences drawn from a concentration or specialization measure will thus be ambiguous unless they are based on a clear-cut definition of what kind of concentration or specialization a researcher actually intends to study, and unless the measure reflects this definition properly. And second, the studies differ in the sectoral and spatial scales of the data used to calculate the measures. They largely neglect the fact that the choice of the sectoral and spatial scales affects, due to the *checkerboard problem and the modifiable areal unit problem* (MAUP),[3] the values of the measures and their interpretation.

---

1 Examples of such inequality measures are the Theil index, the Gini coefficient, the coefficient of variation, and the relative mean deviation (Krugman index). 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), or Nijkamp et al. (2003) for recent reviews.

2 Allison (1978: 865) makes a similar point with respect to the measurement of income inequality.

3 The checkerboard problem (Arbia 2001) arises from neglecting relevant information on the locations of or distances between regions (or industries; see Arbia and Piras 2008). Treating regions as anonymous units renders the measures insensitive, inter alia, to whether or not regions with similar characteristics are spatially clustered. The MAUP (Openshaw and Taylor 1979; Arbia 1989) arises from discretizing heterogeneous continuous space into regions. It comes under two guises: (i) Discretizing space averages away heterogeneity within the regions. The larger the regions are defined, the more information on spatial heterogeneity will

The present paper addresses these two deficiencies in the concentration and specialization literature.[4] It addresses the first deficiency by redefining and extending the available inequality measures such that they are better suited to study concentration and specialization patterns, and by proposing a taxonomy of these measures. The redefined and extended inequality measures will be called *disproportionality* measures of regional concentration or of industrial specialization. The taxonomy of these disproportionality measures helps researchers to *specify* their research hypothesis more precisely, and to *define* a measure that appropriately reflects their definition of concentration or specialization. The paper addresses the second deficiency by generalizing disproportionality measures of concentration and their taxonomy to—what will be called—*spatial* disproportionality measures of concentration (*spatial concentration* measures for short). Spatial concentration measures mitigate the effects of the checkerboard problem and the MAUP by explicitly taking into account an industry's characteristics in neighboring regions.

In addition to addressing the two deficiencies of the literature, the present paper proposes disproportionality measures of economic localization (*localization* measures for short) as well as a taxonomy of these measures. Being straightforward generalizations of the disproportionality measures of concentration and specialization, localization measures overcome the dichotomy between concentration and specialization. Summarizing the joint distribution of, say, employment across regions and industries in a single measure, they permit a simultaneous analysis of regional concentration and industrial specialization patterns, and thereby facilitate a nested analysis of concentration and specialization patterns at different sectoral and regional scales.

The taxonomy of disproportionality measures gives rise to a modular construction system of three characteristic features that unambiguously define any disproportionality measure: a weighting scheme, a reference distribution, and a projection function.[5] Each characteristic

---

usually be lost (*scale* problem). The concentration or specialization measures, which are supposed to measure the extent of spatial heterogeneity, are therefore sensitive to the scale of spatial aggregation. By the same token, they are sensitive to the scale of sectoral aggregation. (ii) The boundaries between the discrete spatial (or sectoral) units may be such that they cut through geographical areas (or sectoral activities) that are homogeneous in terms of their characteristics of interest, or such that the spatial (sectoral) units comprise areas (or activities) with distinctively different characteristics (*arbitrary boundary* problem).

4 A third deficiency in the literature, which is discussed only briefly in Section 2 of this paper, is the lack of rigorous statistical hypothesis testing (Combes and Overman 2004). Taking the values of measures at face value, most studies neglect the fact that the measures may be subject to a significant amount of uncertainty.

5 Spatial disproportionality measures have a fourth characteristic feature, the spatial weighting scheme.

feature serves a distinct function for the measure: The weighting scheme determines the basic units (subjects) of the analysis, the reference distribution the benchmark of no concentration or specialization, and the projection function the relative emphasis put on positive and negative as well as on large and small deviations of the observed units from their reference. For each characteristic feature, there is a variety of alternative realizations that can be chosen largely independently of the realizations of the other two features.[6] By purposefully choosing a specific realization for each feature, a researcher can *define* the disproportionality measure that meets the requirements of his specific research hypothesis most closely, and he can alleviate, as far as possible, the effects of a suboptimal sectoral and spatial disaggregation of his data on his inferences.

The three characteristic features jointly determine the informational content and interpretation of the measure: The value of a measure may thus be attributed to the concept of concentration or specialization implied by the measure's characteristic features. Moreover, the robustness of the inferences drawn from a measure can be investigated by selectively modifying the realizations of single characteristic features. The taxonomy of measures also helps researchers to specify their research hypothesis more precisely by giving them the opportunity to assess the implications for the measure's interpretation of different realizations of each characteristic feature.

The taxonomy of measures 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.[7] It actually covers a much wider range of measures, including the entire generalized entropy (GE) class of measures.[8] 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 using Ripley's K as functions proposed by Duranton and Overman (2005; 2008) and Marcon

6 The bulk of the existing literature considers inequality measures as fixed combinations of two of these three features, the reference distribution and the projection function. An important step towards a more flexible combination of features is taken by Brülhart and Träger (2005), who consider using different references for the same projection function. However, unlike the present paper, they do not consider varying the weighting scheme independently of the reference distribution.

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

8 The GE class of measures on its part is closely linked to the extended Atkinson class of measures (Lasso de la Vega and Urrutia forthcoming), which is, like the GE class, very prominent in the income inequality literature.

and Puech (2003; 2005). Disproportionality measures similar to those discussed in the present paper may also be useful for the analyses of regional income inequalities or international trade patterns because these analyses are subject to similar conceptual problems.

The organization of the paper is as follows. Section 2 introduces the disproportionality measures of concentration and specialization as well as their taxonomy, Section 3 extends the taxonomy to disproportionality measures of localization, and Section 4 extends it to spatial disproportionality measures of concentration. Section 5 concludes, and discusses possible directions for future research. Appendix 1 deals with the issue of standardizing the disproportionality measures discussed in this paper to the 0 – 1 interval by dividing them by their upper bounds. This paper does not give specific empirical illustrations of the measures discussed. Any short and quick empirical illustrations would contradict one of the main purposes of this paper, which is to emphasize the need to define research purposes carefully and unambiguously, and to define measures such that they fit both the research purpose and the characteristics of the available data as closely as possible.[9]

2. The Taxonomy

This section introduces a taxonomy of disproportionality measures of concentration and specialization. It defines the disproportionality measures, discusses their three characteristic features, and illustrates how specific disproportionality measures of concentration and specialization can be constructed using a modular construction system consisting of the three characteristic features. This section also briefly discusses merits and drawbacks of adopting methods of statistical hypothesis testing used in the income inequality literature.

Definition of disproportionality measures

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 empirical studies of concentration or specialization, this section focuses on discrete versions of the measures. The population may be workers, establishments, or units of value added; the char-

---

[9] The interested reader is referred to Bickenbach et al. (2008), which investigates the evolutions of localization, concentration and specialization in Europe, using the measures discussed in the present paper.

acteristics may be industries or regions. For expositional convenience, this section is limited to measures of the regional concentration of employment in an industry. Thus, the population is *workers within an industry*, and their characteristics are the *regions* of their workplaces.[10]

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 $\mathbf{L}_{(ir)} = (L_{ir}\text{: } ir \in \mathbf{I} \times \mathbf{R})$ denote the industry-region employment pattern and $\mathbf{L}_{i(r)} = (L_{ir}\text{: } r \in \mathbf{R})$ the distribution of industry i employment across regions at a given point in time.[11] $\mathbf{L}_{i(r)}$ will henceforth be dubbed the "variable of main interest". For a given reference distribution, $\mathbf{\Pi}_{(r)} = (\Pi_r\text{: } r \in \mathbf{R})$, and absolute region-specific weights, $\mathbf{W}_{(r)} = (W_r\text{: } r \in \mathbf{R})$, the disproportionality measure $M_i^{W\Pi}$ is given by

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

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}], are always scaled by their weighted average across all regions, $\overline{X}_i = \sum_{r=1}^R w_r X_{ir} = \sum_{r=1}^R \frac{W_r}{\sum_r W_r} \frac{L_{ir}}{\Pi_r}$. Likewise, the region-specific weights, W_r, are always scaled by their average across all regions, such that $w_r = W_r/\Sigma_r W_r$. Technically, M_i is a function of w_r and $X_{ir}/\overline{X}_i$ only, similar to inequality measures. The key difference is, however, that a disproportionality measure 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. This—apparently minor—extension or redefinition makes the measures more suitable for analyzing concentration or specialization because it introduces the reference explicitly as a separate variable into the measure. This reference can be chosen from a wide array of possible references, depending on the research purpose at hand. The population mean, $\overline{X}_i$, merely serves as a scaling factor in disproportionality measures. It ensures that the measures assume a minimum value of zero if the region-specific disproportionality 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 measures invariant to the

---

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

[11] The time index, t, is omitted here for simplicity.

scales of both the variable of main interest and the reference. The scale of the reference may thus deviate from that of the variable of main interest.

The three characteristic features

The taxonomy proposed in this paper 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, $\mathbf{L}_{i(r)}$, these 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 analysis (Brülhart and Träger 2005): for measures of concentration, the basic units are spatial units. The variable of main interest is defined as the number of industry i workers *per basic spatial unit*.[12] 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 basic units to be specified but require the variable of main interest as well as the references to 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). By standardizing the sum of weights to one, each region is assigned the relative weight $w_r = 1/R$. Second, square kilometers (km²) have been chosen as basic units, which implies weighting each region by its geographical size (A_r), i.e., $\mathbf{W}_{(r)} = \mathbf{A}_{(r)} = (A_1, \ldots, A_R)$. 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 a region. And third, the average size of the area attributed to a worker in the region in the year, t, under study 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. Types of basic units that have not been used in the litera-

---

12 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 unit of (sectoral) activities.

ture so far include (i) the average size of the area attributed to a worker in a fixed reference year, t_0, i.e., $\mathbf{W}_{(r)t} = \mathbf{L}_{\bullet(r),t0} = (L_{\bullet 1,t0}, \ldots, L_{\bullet R,t0})$, $t \neq t_0$, or (ii) the average size of the area attributed to a worker in the region-industry itself in t_0, i.e., $\mathbf{W}_{(r)t} = \mathbf{L}_{ir,t0} = (L_{i1,t0}, \ldots, L_{iR,t0})$, $t \neq t_0$. These fixed-year weights are, like the uniform weights, time-invariant, a property that is particularly useful for studying changes in the concentrations of large industries or sectors over time. Unlike contemporary weights drawn from higher-level sectoral aggregates, fixed-year weights can be assumed to be exogenous to employment changes in the industry under study even if this industry is large.

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 if the weights reflect the sizes of the regions, and the subregions exhibit, or are assumed to exhibit, identical concentration patterns. Haaland et al. (1998) attribute this invariance to relative measures. The taxonomy proposed here makes clear that it is solely due to the choice of the weights.

(ii) The *reference distribution*, $\mathbf{\Pi}_{(r)}$, reflects the researcher's choice of 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 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 the three types of references that match the weights discussed above have, nonetheless, been used in the literature: the uniform distribution, the distribution of the geographical sizes of regions, or the distribution of employment observed at a higher-level sectoral aggregate. Choosing the uniform distribution as the reference implies assuming all regions to be of the same size under the null hypothesis (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. Choosing the distribution of the geographical sizes of regions as the reference implies assuming employment in the industry under investigation to be distributed evenly across space under the H_0, i.e., $\mathbf{\Pi}_{(r)} = \mathbf{A}_{(r)}$. And choosing the distribution of employment observed at a higher-level sectoral aggregate (total regional employment, for example) as the reference 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., $\mathbf{\Pi}_{(r)} = \mathbf{L}_{\bullet(r)}$. The latter choice reflects the researcher's emphasis on controlling for systematic differences between regions in the sizes of their labor forces, in their attractiveness to firms or workers, their regulatory frameworks, or other institutional or political factors. Types of references that have not been used in the literature so far include (i) the distribution of a higher-level sectoral aggregate in a fixed reference year t_0, i.e., $\mathbf{\Pi}_{(r)t} = \mathbf{L}_{\bullet(r),t0}$, $t \neq t_0$, and (ii) the spatial distribution of employment in the industry under investigation itself in t_0, i.e., $\mathbf{\Pi}_{(r)t} = \mathbf{L}_{i(r),t0}$, $t \neq t_0$. These fixed-year references are time-invariant, like uniform references, but allow controlling for differences in the exogenous sizes of regions. Measures based on the uniform reference will henceforth be labeled *absolute* measures, those based on a nonuniform reference, *relative* measures.[13]

(iii) The *projection function*, f_M, generally reflects the researcher's relative emphasis on region-specific proportionality factors of different magnitudes (Cowell, 2000; Cowell and Flachaire, 2002). Types of projection functions covered by the taxonomy include the generalized entropy (GE) class of measures, the relative mean deviation (RMD), and the Gini coefficient.[14]

According to "folk wisdom" (Cowell and Flachaire, 2002: 1), different projection functions emphasize different values of the region-specific proportionality factors. Projection functions of the GE type with a sensitivity parameter $\alpha < 0$ ($\alpha > 1$) are said to put stronger emphasis on variations in the range of lower (higher) values of the region-specific proportionality factors. The respective projection functions may consequently be preferred if, in a study of the evolution of concentration over time, for example, the test hypothesis suggests putting particular emphasis on the changes in the regions where the industry is *under*represented (*over*represented). The projection function of the RMD is said to put stronger emphasis on changes in the balance between below- and above-average values of the region-specific proportionality factors, and on incidences of regions "jumping across" the reference. This projection function may consequently be preferred if the test hypothesis suggests putting more emphasis on changes in, or differences between, the aggregate balance of *under*- and *over*representation.

---

[13] 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.

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

And the projection function of the Gini coefficient is said to put neither particular emphasis on regions where the industry is strongly *under*represented nor on those where the industry is strongly *over*represented.

In addition to being subject to these theoretical considerations, the choice of the projection function may be subject to practical considerations. To reduce the effects of "outliers" or indivisibilities in firm sizes on the measure, a projection function may be preferred that is not too sensitive to these outliers or indivisibilities. The "folk wisdom" would, for example, suggest preferring a GE measure with $\alpha > 1$ ($\alpha < 0$) if the data is contaminated by outlying regions with low (high) proportionality factors (Cowell and Flachaire 2002). It also suggests that the RMD and the Gini coefficient are not too sensitive to both regions with particularly low and those with particularly high proportionality factors.

To assess the robustness of the results obtained for the preferred projection function, they may be compared to the results obtained for other projection functions. The taxonomy proposed here offers the most suitable framework for these robustness tests because it facilitates changing the projection function while retaining the preferred region-specific weights and references.

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 three 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 comparing 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, the research purpose or test hypothesis may require comparing the values of two measures that differ only in the reference or in the weighting scheme. And third, selectively changing the region-specific weights or the references helps in assessing the robustness of the preferred measure to a variation of the basic units or the null hypothesis of no concentration.

Illustrations

Using the taxonomy proposed here, 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 gives an overview of the general principle of defining disproportionality measures of concentration for selected projection functions: the GE class of measures, the Theil index, the CV,[15] the *RMD*, and the Gini coefficient. The table can easily be extended to projection functions based on other measures discussed in the inequality literature (see, e.g., Cowell 1995; Silber 1999). The first column in 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 in 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] To save space, weighted absolute measures are omitted, and the relative and the weighted measures are exemplified only for total regional employment as references or weights.

In what follows, the constructive principle of disproportionality measures will be illustrated for two projection functions: the relative mean deviation (RMD; "Krugman" index), and the generalized entropy (GE) class of measures. The illustration of the RMD/Krugman index

---

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.

16 All three variants of the measures listed in Table 1 have actually been employed in studies of regional concentration or industrial 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), Traistaru et al. (2003), Morgenroth (2008), and Totev (2008); (ii) the so-called relative Theil index (e.g., Brülhart and Träger 2005, Iara 2008, Krieger-Boden 2008a; 2008b), (iii) the relative CV (e.g., Brülhart and Träger 2005), and (iv) the "locational" Gini coefficient (Krugman 1991, Amiti 1998, Brülhart 2001). An *unweighted relative* measure is the Gini coefficient used by Südekum (2006), who (mis-) interprets his Gini coefficient as the locational Gini. And among the *unweighted absolute* measures are (i) the traditional Gini coefficient (Aiginger and Leitner 2002, Midelfart-Knarvik et al. 2002), (ii) the absolute Theil index (Aiginger and Davies 2004), and (iii) the absolute CV (Aiginger and Leitner 2002).

Table 1. Disproportionality Measures of Regional Concentration

Projection Function	General Form	Unweighted Absolute	Unweighted Relative	Weighted Relative
		References: $\Pi_r := 1$ Weights: $W_r := 1 \Rightarrow w_r = 1/R$	References: $\Pi_r := L_{\bullet r}$ Weights: $W_r := 1 \Rightarrow w_r = 1/R$	References: $\Pi_r := L_{\bullet r}$ Weights: $W_r := L_{\bullet r} \Rightarrow w_r = \lambda_r$
$\mathrm{GE}_{(\alpha)}$	$GE_{(\alpha)i} = \omega \sum_{r=1}^{R} w_r \left[\left(\frac{\frac{L_{ir}}{\Pi_r}}{\sum_r w_r \frac{L_{ir}}{\Pi_r}} \right)^{\alpha} - 1 \right]$	$GE_{(\alpha)i}^{UA} = \omega \left[\frac{1}{R} \sum_{r=1}^{R} (R\lambda_{ir})^{\alpha} - 1 \right]$	$GE_{(\alpha)i}^{UR} = \omega \left[\frac{1}{R} \sum_{r=1}^{R} \left(R \frac{l_{ir}}{\sum_r l_{ir}} \right)^{\alpha} - 1 \right]$	$GE_{(\alpha)i}^{WR} = \omega \left[\sum_{r=1}^{R} \lambda_r (LC_{ir})^{\alpha} - 1 \right]$
Theil $= \mathrm{GE}_{(1)}$	$T_i = \sum_{r=1}^{R} w_r \frac{\frac{L_{ir}}{\Pi_r}}{\sum_r w_r \frac{L_{ir}}{\Pi_r}} \ln \left(\frac{\frac{L_{ir}}{\Pi_r}}{\sum_r w_r \frac{L_{ir}}{\Pi_r}} \right)$	$T_i^{UA} = \sum_{r=1}^{R} \lambda_{ir} \ln(R\lambda_{ir})$	$T_i^{UR} = \sum_{r=1}^{R} \frac{l_{ir}}{\sum_r l_{ir}} \ln \left(R \frac{l_{ir}}{\sum_r l_{ir}} \right)$	$T_i^{WR} = \sum_{r=1}^{R} \lambda_{ir} \ln(LC_{ir})$
CV = $(2\mathrm{GE}_{(2)})^{0.5}$	$CV_i = \left[\sum_{r=1}^{R} w_r \left[\left(\frac{\frac{L_{ir}}{\Pi_r}}{\sum_r w_r \frac{L_{ir}}{\Pi_r}} \right)^{2} - 1 \right] \right]^{0.5}$	$CV_i^{UA} = \left(R \sum_{r=1}^{R} \lambda_{ir}^2 - 1 \right)^{0.5}$	$CV_i^{UR} = \left[R \sum_{r=1}^{R} \left(\frac{l_{ir}}{\sum_r l_{ir}} \right)^2 - 1 \right]^{0.5}$	$CV_i^{WR} = \left(\sum_{r=1}^{R} \lambda_r LC_{ir}^2 - 1 \right)^{0.5}$
RMD / Krugman	$RMD_i = \frac{\sum_{r=1}^{R} w_r \left\| \frac{L_{ir}}{\Pi_r} - \sum_r w_r \frac{L_{ir}}{\Pi_r} \right\|}{\sum_r w_r \frac{L_{ir}}{\Pi_r}}$	$RMD_i^{UA} = \sum_{r=1}^{R} \left\| \lambda_{ir} - \frac{1}{R} \right\|$	$RMD_i^{UR} = \sum_{r=1}^{R} \left\| \frac{l_{ir}}{\sum_r l_{ir}} - \frac{1}{R} \right\|$	$RMD_i^{WR} = \sum_{r=1}^{R} \left\| \lambda_{ir} - \lambda_r \right\|$
Gini	$G_i = \frac{\sum_{r=1}^{R} \sum_{s=1}^{R} w_r w_s \left\| \frac{L_{ir}}{\Pi_r} - \frac{L_{is}}{\Pi_s} \right\|}{2 \sum_r w_r \frac{L_{ir}}{\Pi_r}}$	$G_i^{UA} = \frac{1}{2R} \sum_{r=1}^{R} \sum_{s=1}^{R} \left\| \lambda_{ir} - \lambda_{is} \right\|$	$G_i^{UR} = \frac{1}{2R \sum_r l_{ir}} \sum_{r=1}^{R} \sum_{s=1}^{R} \left\| l_{ir} - l_{is} \right\|$	$G_i^{WR} = \frac{1}{2} \sum_{r=1}^{R} \sum_{s=1}^{R} \lambda_r \lambda_s \left\| LC_{ir} - LC_{is} \right\|$

Notation: $L_{i\bullet} := \Sigma_r L_{ir}$; $L_{\bullet r} := \Sigma_i L_{ir}$; $l_{ir} := L_{ir}/L_{\bullet r}$; $l_i := L_{i\bullet}/\Sigma_i L_{i\bullet}$; $\lambda_{ir} := L_{ir}/L_{i\bullet}$; $\lambda_r := L_{\bullet r}/\Sigma_r L_{\bullet r}$; $LC_{ir} := l_{ir}/ l_i = \lambda_{ir}/\lambda_r$; $\omega := (\alpha^2\text{-}\alpha)^{-1}$.

demonstrates that even measures that are usually considered ad-hoc measures fit into the taxonomy after a suitable transformation and reinterpretation.[17] The illustration of the GE class of measures demonstrates the interpretation of disproportionality measures in terms of basic units.

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 between the 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 correspond exactly to the reference distribution. The Krugman index has traditionally been classified as a "relative" measure. 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):

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

$LC_{ir} = \lambda_{ir}/\lambda_r$ in (3) denotes the location coefficient for region-industry ir. 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 in Table 1.

The first line of (3) illustrates the constructive principle of all the disproportionality measures discussed in the present paper: Any disproportionality measure first determines the (dis-) pro-

[17] Essentially, the same is true for the Gini coefficient in general and the locational Gini coefficient in particular. The constructive principle of Gini disproportionality measures and the corresponding Lorenz curves is illustrated in Appendix 2.

portionality factor for each region by comparing the value for the region-industry, L_{ir}, to its corresponding reference value, which is $L_{\bullet r}$ in this case. And second, the measure converts the region-specific proportionality factors into its specific metric by applying the projection function. The projection function of the *RMD* requires (i) scaling the region-specific proportionality factors by their weighted mean across all regions, $\Sigma_r(L_{\bullet r}/L_{\bullet\bullet})(L_{ir}/L_{\bullet r})$ [=: l_i], employing the weights implied by the choice of the basic units; (ii) subtracting 1; (iii) taking the absolute value; and (iv) taking the weighted average over all regions, again employing the weights implied by the choice of the basic units.

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

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

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

$$RMD_i^{UA} = f_{RMD}\left(\mathbf{1}_{(r)}, \frac{\mathbf{L}_{i(r)}}{\mathbf{1}_{(r)}}\right) = \sum_{r=1}^{R}\left|\lambda_{ir} - \frac{1}{R}\right| \tag{5}$$

(see Table 1, second column). Comparing the Krugman index, or weighted relative RMD in (2) to the unweighted absolute RMD in (5) demonstrates the usefulness of the proposed 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 their reference. The proposed taxonomy shows that the two measures actually differ in two characteristics, their reference *and* their region-specific weights.

Consider second the generalized entropy class of measures, $GE_{(\alpha)}$. In contrast to the *RMD*, the GE measures can be derived from a set of axioms that define several useful properties (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 the inequality between any set of subgroups of the population. The GE class of measures for a vector of characteristics $\mathbf{Y}_{(n)} = (Y_1, \ldots, Y_N)$ of a population of N "basic units" is generally defined as

$$GE_{(\alpha)} = f_{GE(\alpha)}\left(\mathbf{Y}_{(n)}\right) = \frac{1}{\alpha(\alpha-1)}\frac{1}{N}\sum_{n=1}^{N}\left[\left(\frac{Y_n}{\bar{Y}}\right)^{\alpha} - 1\right], \qquad \alpha \neq 0, 1, \tag{6}$$

with the corresponding limiting expressions for $\alpha = 0$ and $\alpha = 1$. $\overline{Y} = \frac{1}{N}\sum_n Y_n$ denotes the mean across all members of the population. The most prominent GE measures are those given by $\alpha = 2$, which is a simple monotonic transformation of the coefficient of variation, $GE_{(2)} = \frac{1}{2}CV^2$, and by $\alpha \to 1$, which is the Theil index, i.e., $GE_{(1)} = T$.

$GE_{(\alpha)}$ in (6) can be decomposed into a within-groups component, $GE_{(\alpha)w}$, 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

$$GE_{(\alpha)b} = f_{GE(\alpha)}\left(\overline{Y}_{(h)}, \mathbf{N}_{(h)}\right) = \frac{1}{\alpha(\alpha-1)}\sum_{h=1}^{H}\frac{N_h}{N}\left[\left(\frac{\overline{Y}_h}{\overline{Y}}\right)^{\alpha} - 1\right], \tag{7}$$

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

Traditionally, the (unweighted) absolute $GE_{(\alpha)}$ measures of concentration have been derived from (6) and the (weighted) relative measures from the between-group component (7), 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 (7) as the unique basis for all GE measures of concentration (see Table 1, first row):

$$GE_{(\alpha)i} = f_{GE(\alpha)}\left(\mathbf{W}_{(r)}, \frac{\mathbf{L}_{i(r)}}{\mathbf{\Pi}_{(r)}}\right) = \frac{1}{\alpha(\alpha-1)}\sum_{r=1}^{R} w_r\left[\left(\frac{\dfrac{L_{ir}}{\Pi_r}}{\sum_r w_r \dfrac{L_{ir}}{\Pi_r}}\right)^{\alpha} - 1\right]. \tag{8}$$

All disproportionality measures derived from (8) share the usual properties of GE measures, provided the variables of main interest, the weights, and the references are related to the basic units in a consistent way.

Statistical inferences

Since the data used in studies of concentration or specialization usually comprises the entire populations of regions, industries, workers, or units of value added rather than nonexhaustive random samples of a larger population, the measures calculated from this data may be considered non-stochastic. This is not the case, however, if the data on region-industry employment or value added is subject to measurement errors or idiosyncratic random shocks, which introduces a stochastic component to the measures (Brülhart and Träger 2005).

Those studies that account for this stochastic component by assessing the statistical significance of observed concentration or specialization patterns use inferential methods borrowed from the income inequality literature. In the income inequality literature, two approaches have been used to calculate standard errors or probability intervals: asymptotic theory (Cowell 1989) and bootstrap methods (Mills and Zandvakili 1997; Biewen 2002). In this literature, the main source of uncertainty is that the observed income data is generated by a nonexhaustive random sampling process. Although the bootstrap method can be expected to perform better than asymptotic theory in income inequality studies (Cowell and Flachaire 2002), both methods are not without problems. One problem is that the theoretical properties of these methods are largely unknown for small samples. Another problem is that the errors in rejection probabilities can be significant even for very large samples, as simulations show. These errors tend to be quite sensitive to the characteristics of the distribution of the observed data (Cowell and Flachaire 2002).

The inferential methods used for income inequality measures can, in principle, also be used for the corresponding disproportionality measures discussed in the present paper. They may help in evaluating the significance of changes in concentration (or specialization) over time, or that of differences across industries or regions. Brülhart and Träger (2005), for example, employ a (weighted block-) bootstrap based on Biewen (2002) to assess the significance of changes in the regional concentration of industries over time. This test is motivated by the assumption that their data is subject to country-specific measurement errors, and that these errors are independent of the sizes of the region-industries in terms of employment or value added.

Any statistical tests for measures of concentration or specialization based on these methods should, however, be interpreted with even greater caution than those for measures of income inequality for at least two reasons. First, small sample errors may be more relevant because the number of observations in concentration or specialization studies is typically much smaller. And second, insights on the performance of these tests in the context of income inequality measurement are of only limited relevance because they are gained for the specific sources of uncertainty and the specific characteristics of distributions usually observed for income data. These sources of uncertainty and characteristics of distributions may differ fundamentally from those considered relevant in the analysis of concentration or specialization. While these differences are, under appropriate regularity assumptions, irrelevant for the *asymptotic* properties of the statistical tests, they may affect the properties of the statistical tests significantly for finite samples of regions or industries. For finite samples, the reliability of the tests appears to be generally quite sensitive to the specific sources of uncertainty and characteristics of distributions. As a consequence, bootstrap tests of the equality of two values of a measure may be far too conservative, rejecting the null hypothesis of equality in far too few cases, if the sample sizes and the measurement errors are rather small. Detailed studies are warranted to assess the performances of the statistical tests in the measurement of concentration or specialization for relevant distributional assumptions and sample sizes.

3. Generalization 1: Disproportionality Measures of Localization

Disproportionality measures of localization 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 localization cover both dimensions simultaneously, thus evaluating all elements of the industry-region employment pattern, $\mathbf{L}_{(ir)}$.[18] In terms of the proposed taxonomy, localization measures require specifying an (IxR) matrix of two-dimensional *region-industry*-specific weights, $\mathbf{W}_{(ir)}$ [= (W_{ir}: $i \in \mathbf{I}$; $r \in \mathbf{R}$)], and an (IxR) matrix of two-dimensional *region-industry*-specific references $\mathbf{\Pi}_{(ir)}$. The region-industry-specific weights reflect the choice of basic units in the regional and sectoral dimensions; the references reflect the no-localization benchmark for each region-industry under study. The weights and references may—but need not necessarily—be products of industry-specific and region-specific weights and references, for example, $\mathbf{W}_{(ir)} = \mathbf{W}_{(i)}\mathbf{W}_{(r)}$\` or $\mathbf{\Pi}_{(ir)} = \mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}$\`.

---

[18] Similar measures for two-dimensional data, labeled “segregation measures”, have recently been discussed in the sociological segregation literature (Reardon and Firebaugh 2002). In the economics literature, special cases of the localization measures discussed here have been used by Aiginger and Davies (2004), Aiginger and Rossi-Hansberg (2006), Cutrini (2006), and Mulligan and Schmidt (2005). The present paper generalizes these measures in order to integrate them into the taxonomy presented in Section 2.

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

$$GE_{(\alpha)}^{WR} = f_{GE(\alpha)}\left(\mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}\text{\`}, \frac{\mathbf{L}_{i(r)}}{\mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}\text{\`}}\right) = \frac{1}{\alpha(\alpha-1)}\sum_{i=1}^{I}\sum_{r=1}^{R} l_i \lambda_r \left(LC_{ir}^{\alpha} - 1\right). \tag{9}$$

For $I = 1$, all localization measures in Table 2 reduce to the corresponding concentration measures discussed in Section 2; for $R = 1$, they reduce to the corresponding specialization measures.

Notice that localization measures do not require specifying the same kinds of weights or references in the sectoral and regional dimensions. A research hypothesis that suggests, for example, choosing relative weights in the sectoral dimension but absolute weights in the regional dimension can be investigated by specifying a region-industry-specific weights matrix such as $\mathbf{W}_{(ir)} = \mathbf{L}_{(i)\bullet}\mathbf{1}_{(ir)}$\`. Each region-industry is assigned the standardized weight $w_{ir} = W_{ir}/(\Sigma_i\Sigma_r W_{ir}) = l_{i\bullet}/R$ in this case. Similarly, the references can be constructed from combinations of different types of industry- and region-specific references. Choosing total employment by industry as a sectoral and area as a regional reference implies, for example, the reference matrix $\mathbf{\Pi}_{(ir)} = \mathbf{L}_{(i)\bullet}\mathbf{A}_{(r)}$\`. The definitions of the references or weights may even differ between the industries or the regions as long as the scales of the references (or weights) are the same for all region-industries. For example, the regional reference (and/or weight) for agriculture may be chosen to be a region's area, for manufacturing industries the region's aggregate employment, and for service industries the region's population.

Localization measures share virtually all the properties of the corresponding concentration or specialization measures (Reardon and Firebaugh 2002). One of these properties, the decomposability of the GE measures, is particularly useful for testing detailed and complex research hypotheses about the contributions of individual industries, sectors, counties, or states within a nation to the national economy's overall localization at a specific point in time. GE measures of localization can be decomposed successively or simultaneously in the industrial and the regional dimensions. Appendix 3 illustrates the stepwise decomposition of a Theil index

Table 2. Disproportionality Measures of Economic Localization[1]

Measure	General Form
$GE_{(\alpha)}$	$GE_{(\alpha)} = \frac{1}{\alpha(\alpha-1)} \sum_{i=1}^{I}\sum_{r=1}^{R} w_{ir} \left[\left(\frac{\frac{L_{ir}}{\Pi_{ir}}}{\sum_{i,r} w_{ir} \frac{L_{ir}}{\Pi_{ir}}} \right)^{\alpha} - 1 \right]$
Theil = $GE_{(1)}$	$T = \sum_{i=1}^{I}\sum_{r=1}^{R} w_{ir} \frac{\frac{L_{ir}}{\Pi_{ir}}}{\sum_{i,r} w_{ir} \frac{L_{ir}}{\Pi_{ir}}} \ln \left(\frac{\frac{L_{ir}}{\Pi_{ir}}}{\sum_{i,r} w_{ir} \frac{L_{ir}}{\Pi_{ir}}} \right)$
CV = $(2GE_{(2)})^{0.5}$	$CV = \left[\sum_{i=1}^{I}\sum_{r=1}^{R} w_{ir} \left[\left(\frac{\frac{L_{ir}}{\Pi_{ir}}}{\sum_{i,r} w_{ir} \frac{L_{ir}}{\Pi_{ir}}} \right)^{2} - 1 \right] \right]^{0.5}$
RMD / Krugman	$RMD = \sum_{i=1}^{I}\sum_{r=1}^{R} w_{ir} \left| \frac{\frac{L_{ir}}{\Pi_{ir}}}{\sum_{i,r} w_{ir} \frac{L_{ir}}{\Pi_{ir}}} - 1 \right|$
Gini	$G = \frac{1}{2\sum_{i,r} w_{ir} \frac{L_{ir}}{\Pi_{ir}}} \sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{r=1}^{R}\sum_{s=1}^{R} w_{ir} w_{js} \left| \frac{L_{ir}}{\Pi_{ir}} - \frac{L_{js}}{\Pi_{js}} \right|$

1 The corresponding unweighted absolute, unweighted relative, and weighted relative measures are obtained from the general forms in the same way as described for regional concentration measures in Section 2 and Table 1.

of a country's localization by sectors, states, and industries. A comparison over time of the measures obtained by decompositions can be used to test research hypotheses about the contributions of individual industries, sectors, counties, or states to changes in overall national localization (Bickenbach et al. 2008).

If the (contemporary) industry and region totals are used as references and weights in a weighted relative localization measure, i.e., if $\mathbf{\Pi}_{(ir)} = \mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}`$ and $\mathbf{W}_{(ir)} = \mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}`$, the localization measure is simply the weighted average 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 requirements. For the RMD, for example, one obtains

$$
\begin{aligned}
RMD^{WR} = f_{MRD}\left(\mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}`, \frac{\mathbf{L}_{(ir)}}{\mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}`}\right) &= \sum_{i=1}^{I}\sum_{r=1}^{R} l_i \lambda_r \left|LC_{ir} - 1\right| \\
&= \sum_{i=1}^{I} l_i RMD_i^{WR} = \sum_{r=1}^{R} \lambda_r RMD_r^{WR}.
\end{aligned} \tag{10}
$$

4. Generalization 2: Spatial Disproportionality Measures of Concentration

All the disproportionality measures discussed so far are invariant to the spatial ordering of, and the interdependencies between the regions under investigation, which gives rise to the checkerboard problem and the MAUP (see footnote 3). 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. While the inequality measure is informative as to the degree of concentration of an industry within regions, the spatial statistic gives an indication of the degree of spatial clustering of the industry between regions. Lafourcade and Mion (2007) use a similar approach 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 such 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 values of the variable of interest and the reference of any region with the corresponding values of nearby regions. Reardon and O'Sullivan (2004) suggest doing so in a way similar to a kernel density estimation, or a geographically weighted analysis. More specifically, they suggest defining a measure in terms of the spatially weighted averages of the variables of main interest and the reference.[19]

To extend the taxonomy presented in Section 2 to spatial disproportionality measures, L_{ir} and Π_r can be redefined as spatially weighted sums,

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

[19] 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 focuses on disproportionality measures for regional aggregates.

where ϕ_{rq} is a nonnegative spatial weight, or spatial discount factor that reflects the "closeness" of region q to region r, and the superscript S denotes spatially weighted variables. The closeness between regions may generally depend on geographic distances, neighborhood patterns, or accessibility. The variables L_{ir}^S and Π_r^S are spatial averages if the spatial weights are normalized by dividing them by their row sums, or spatial sums if the spatial weights are not row-normalized.

All the measures in Table 1 can be extended to spatial measures of concentration by extending the set of characteristic features of the concentration measures discussed in section 2 by an (RxR) matrix $\mathbf{\Phi}_{(r)} = (\phi_{rq}: r, q \in \mathbf{R}; \phi_{rq} \geq 0)$, and by substituting L_{ir}^S and Π_r^S for L_{ir} and Π_r. Table 3 depicts the general forms of the spatial measures for several projection functions, similar to the first column in Table 1. The various weighted or unweighted and absolute or relative measures can be derived from these general forms in a way similar to that outlined in Section 2. The general form of the spatial GE measures, for example, reads

$$GE_{(\alpha)i}^S = f_{GE(\alpha)}\left(\mathbf{W}_{(r)}, \frac{\mathbf{\Phi}_{(r)}\mathbf{L}_{i(r)}}{\mathbf{\Phi}_{(r)}\mathbf{\Pi}_{(r)}}\right) = \frac{1}{\alpha(\alpha-1)}\sum_{r=1}^{R} w_r \left[\left(\frac{\dfrac{\sum_{q=1}^R \phi_{rq}L_{iq}}{\sum_{q=1}^R \phi_{rq}\Pi_q}}{\sum_{r=1}^R w_r \dfrac{\sum_{q=1}^R \phi_{rq}L_{iq}}{\sum_{q=1}^R \phi_{rq}\Pi_q}}\right)^{\alpha} - 1\right]. \quad (11)$$

Due to the geographical weighting, the effect of region r on the measure is larger if industry i is overrepresented (or underrepresented) in both the region itself and its neighbors. And it is smaller if industry i is over- (under-) represented in region r but under- (over-) represented in nearby regions. Setting $\phi_{rq} = 1$ for $q = r$ and $\phi_{rq} = 0$ for $q \neq r$ yields the corresponding aspatial measures.

Spatial concentration measures are capable of mitigating the checkerboard problem and the MAUP inherent to any analysis of concentration based on regional aggregates. They mitigate the checkerboard problem by taking into account the geographical ordering of the regions. The value of a spatial concentration measure will, ceteris paribus, be higher the more the regions where the industry under study is overrepresented or underrepresented are clustered spatially. And they mitigate the MAUP by way of geographical smoothing and carefully specifying the intra- and interregional weights. Geographical smoothing addresses the arbitrary boundary problem. The contribution of a region to the concentration measure takes into account the concentration of the industry on both sides of the region's boundary. And the

Table 3. Spatial Disproportionality Measures of Concentration[1]

Measure	General Form
$GE_{(\alpha)}$	$GE_{i(\alpha)}^{S} = \frac{1}{\alpha(\alpha-1)}\sum_{r=1}^{R} w_r \left[\left(\frac{\frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}}{\sum_{r=1}^{R} w_r \frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}}\right)^{\alpha} - 1\right]$
Theil = $GE_{(1)}$	$T_i^S = \sum_{r=1}^{R} w_r \frac{\frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}}{\sum_{r=1}^{R} w_r \frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}} \ln\left(\frac{\frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}}{\sum_{r=1}^{R} w_r \frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}}\right)$
CV = $(2GE_{(2)})^{0.5}$	$CV_i^S = \left[\sum_{r=1}^{R} w_r \left[\left(\frac{\frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}}{\sum_{r=1}^{R} w_r \frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}}\right)^{2} - 1\right]\right]^{0.5}$
RMD / Krugman	$RMD_i^S = \frac{1}{\sum_{r=1}^{R} w_r \frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}} \sum_{r=1}^{R} w_r \left\lvert \frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q} - \sum_{r=1}^{R} w_r \frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}\right\rvert$
Gini	$G_i^S = \frac{1}{2\sum_{r=1}^{R} w_r \frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q}} \sum_{r=1}^{R}\sum_{s=1}^{R} w_r w_s \left\lvert \frac{\sum_{q=1}^{R}\phi_{rq}L_{iq}}{\sum_{q=1}^{R}\phi_{rq}\Pi_q} - \frac{\sum_{q=1}^{R}\phi_{sq}L_{iq}}{\sum_{q=1}^{R}\phi_{sq}\Pi_q}\right\rvert$

[1] The corresponding unweighted absolute, unweighted relative, and weighted relative spatial measures are obtained from the general forms in the same way as described for regional concentration measures in Section 2 and Table 1.

specification of the intra- and interregional weights addresses the scale problem. Regions are actually treated as being parts of—usually overlapping—larger spatial units that comprise several regions. The intra- and interregional weights determine both the sizes of these larger spatial units, and the relative emphasis put on each of their member regions.

Similar to the choice of the region-specific weights, reference, and projection function, the choice of the spatial weights is subject to the underlying research hypothesis, and may depend on the specificities of the industry under study. If distance-sensitive interactions like trade, commuting, or intra-industry spillovers are considered the main determinants of the regional interdependencies, the geographical weights may be operationalized by some functions of the geographical or economic distances between any two regions, i.e., by $\phi_{rq} = \phi(D_{rq})$, where D_{rq} denotes the distance between the regions q and r, and $\partial\phi/\partial D_{rq} < 0$. Examples of such distance-related spatial weight functions are the exponential distance decay function in which $\phi_{rq} = \exp(-\delta D_{rq})$, and the linear distance decay function in which $\phi_{rq} = (D^* - D_{rq})/D^*$ for $D_{rq} \leq D^*$ and $\phi_{rq} = 0$ for $D_{rq} > D^*$. $\delta\,(\delta > 0)$ denotes the distance decay parameter, which reflects the percentage of spatial loss per unit of distance, and D^* the threshold distance at which regional interdependencies are assumed to approach zero. Alternatively, geographical weights may be operationalized by a variety of other forms of spatial weights (see, e.g., Anselin 1988), including weights based on the existence or the length of a common border, or on the k-nearest neighbors principle.

One possible way to specify 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 merely 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.

It should be noted that, due to geographical smoothing, spatial concentration measures usually assume lower values than their aspatial counterparts. More specifically, for given $\mathbf{L}_{i(r)}$, $\mathbf{\Pi}_{(r)}$, and $\mathbf{W}_{(r)}$, spatial concentration measures assume a minimum value of zero for uniform spatial weights, i.e., for $\phi_{rq} = 1\ \forall\ r, q = 1, \ldots, R$, and a maximum value equal to that of the corresponding aspatial measure for zero interregional weights, i.e., for $\phi_{rq} = 0$ for $q \neq r$ and $\phi_{rq} = 1$ for $q = r$. Moreover, decomposing the spatial GE measures in the usual way is not possible due to the geographical smoothing (Reardon and O'Sullivan, 2004). The regional interdependencies introduced by the geographical weights usually do not allow the set of regions under study to be divided into a smaller set of geographical units such that regional interdependencies do not extend beyond the boundaries of these geographical units.

For georeferenced microdata that provide information on the distances between any pair of establishments, Duranton and Overman (2005; 2008) and Marcon and Puech (2003; 2005) propose describing concentration using functions based on the Ripley's K function. These K-based functions, which assign each possible distance a frequency of observations, 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 to analyze 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.

5. Conclusion

This paper improves and extends the methodological toolbox for analyzing the regional concentration of industries and the industrial specialization of regions. First, it proposes a taxonomy of disproportionality measures of concentration and specialization. These disproportionality 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 using three characteristic features: the weighting scheme, the reference distribution, and the projection function. Each feature can be determined largely independently of the other two features. The modular construction 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 proposed taxonomy to disproportionality measures of economic localization, and to spatial disproportionality measures of concentration. Measures of localization evaluate concentration and specialization patterns simultaneously, and, through decomposition, render possible a nested analysis of the localization, 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 sums or 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 of disproportionality measures proposed in this paper will prove useful for a wide range of empirical studies on concentration, specialization, and localization. It should also prove useful for empirical studies of other economic issues that are subject to similar conceptual problems, such as studies of regional income inequalities or international trade patterns. 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 localization 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 localization measures may be extended to spatially and sectorally weighted localization measures that account for the checkerboard problem and the MAUP in both dimensions. 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. Finally, the reliability of statistical tests for assessing the significance of changes of measures over time, or of the differences between the measures for two regions or industries, needs to be explored in detail.

Appendix 1: Upper Bounds and Standardization of Disproportionality Measures

All the disproportionality measures discussed in this paper have lower bounds of zero, which they take under perfect proportionality. The measures do differ, however, with respect to their upper bounds, which may, but need not, exist.[20] All disproportionality measures for which an upper bound exists can be normalized to vary between 0 and 1 by dividing the measures by their respective upper bounds.

Disproportionality measures of concentration (see Section 2) take their maximum value (if such a value exists) if the industry under study is maximally concentrated. For given weights and references, the industry is maximally concentrated if all employment in the industry is concentrated in the smallest region in terms of basic units, such that $L_{ia} = L_{i\bullet}$ and $L_{ir} = 0$ for $r \neq a,$ where a indexes the smallest region. As a consequence, the upper bound just depends on the smallest region-specific weight, $w_a = \min_r w_r$, as well as the projection function, but not on the references. For the GE measure in equation (8), the upper bound is given by $GE_{(\alpha)i,UB} = [\alpha(\alpha-1)]^{-1}\left(w_a^{1-\alpha}-1\right)$ for $\alpha > 0$. For the other disproportionality measures of concentration given in Table 1 the upper bounds are given by $T_{i,UB} = \ln(1/w_a)$, $CV_{i,UB} = ((1/w_a) - 1)^{0.5}$, $RMD_{i,UB} = 2(1-w_a)$, and $G_{i,UB} = (1-w_a)$. Note that for unweighted measures, all regions are of equal size in terms of basic units, so that w_a=1/R.

Disproportionality measures of localization (see Section 3) take, for given weights, references, and projection function, their maximum value if all employment is clustered in the smallest region-industry (indexed by ba) in terms of basic units such that $L_{ir} = L_{\bullet\bullet}$ for $i = b \wedge r = a$, and $L_{ir} = 0$ otherwise. As a consequence, their upper bound depends only on the smallest region-industry-specific weight, $w_{ba} = \min_{ir} w_{ir}$, as well as the projection function, but not on the references. The upper bound of a $GE_{(\alpha)}$ measure of localization with $\alpha > 0$ is, for example, given by $GE^{WTI}_{(\alpha)UB} = [\alpha(\alpha-1)]^{-1}\left(w_{ba}^{1-\alpha}-1\right)$.[21]

Spatial disproportionality measures of concentration (see Section 4) take their maximum value if the industry is concentrated in a single region, similar to their aspatial counterparts. Calculating the upper bounds of spatial concentration measures is, however, somewhat more

---

[20] An upper bound does not exist for GE measures with $\alpha < 0$.

[21] Again, the GE measure is unbounded for $\alpha < 0$. The upper bounds for the other projection functions in Table 2 can be determined analogously to GE with $\alpha > 0$.

tedious because these upper bounds usually depend on all region-specific weights, the whole reference distribution and all spatial weights. For given region-specific weights, references and spatial weights the upper bound of the spatial $GE_{(\alpha)}$ measure (11) with $\alpha > 0$ is, for example, obtained by solving

$$GE^{S}_{(\alpha)i,UB} = \max_{a} \frac{1}{\alpha(\alpha-1)} \sum_{r=1}^{R} w_r \left[\left(\frac{\frac{\phi_{ra}}{\sum_{q=1}^{R} \phi_{rq} \Pi_q}}{\sum_r w_r \frac{\phi_{ra}}{\sum_{q=1}^{R} \phi_{rq} \Pi_q}} \right)^{\alpha} - 1 \right],$$

where a indexes the region hosting all industry i employment under maximal concentration.[22]

Working with standardized disproportionality measures may have both advantages and disadvantages. On the one hand, working with standardized measures may facilitate comparisons across different spatial or sectoral units, or different spatial or sectoral scales. For unweighted measures of concentration, for example, where the upper bound depends on the number of regions, one may argue that a comparison of the concentration of a particular industry across countries with different numbers of regions is more meaningful when using standardized measures that "control" for the number of regions.[23] On the other hand, changes in standardized measures over time may be largely determined by changes in their upper bounds. This may be particularly problematic for weighted specialization measures if the size of the smallest industry changes significantly and sometimes quite erratically over time, as may easily happen with sectorally highly disaggregated data.[24]

Appendix 2: Defining Gini disproportionality measures

Like 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. The Gini coefficient is generally defined as two times the area between the Lorenz curve and the 45° line (shaded area in Figure A1) in a box plot of cumulated shares of indi-

[22] Generally, the region that solves the maximization problem has to be determined by way of simulation.

[23] One may also argue, however, that the dependence of nonstandardized measures on the number of regions is actually desirable, since an industry that is concentrated in just one region may intuitively be considered less concentrated in a country that has just two regions than in a country that has many regions.

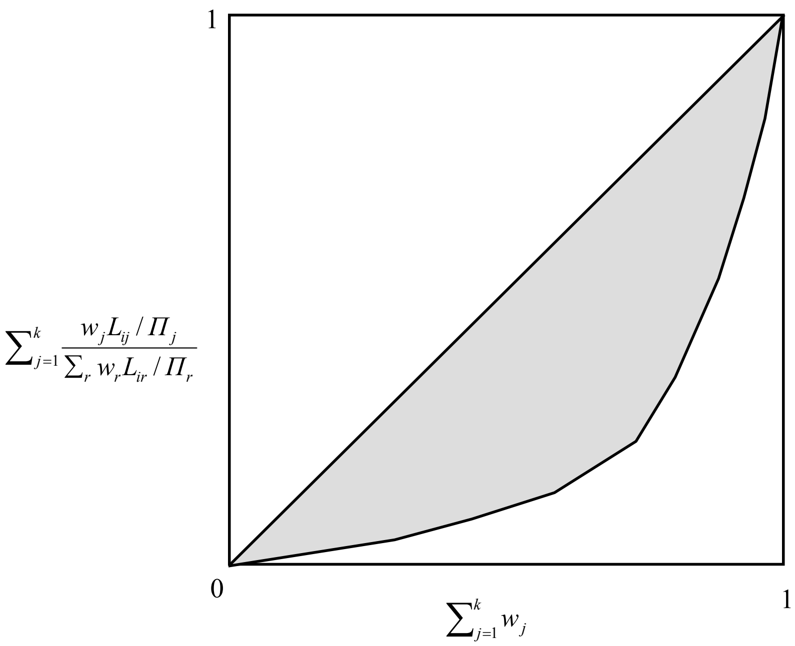
viduals in the population on the horizontal axis and the cumulated shares of their characteristics on the vertical axis.

In terms of the taxonomy of the present paper, the population, depicted on the horizontal axis, consists of the basic units, whose shares are represented by the (relative) region-specific weights, w_r. The characteristics, depicted on the vertical axis, are the weighted region-specific proportionality factors, whose shares are represented by $w_r \frac{L_{ir}}{\Pi_r} \Big/ (\sum_r w_r \frac{L_{ir}}{\Pi_r})$. All observations are sorted in ascending order by the region-specific proportionality factors, L_{ir}/Π_r. This convention gives rise to

$$G_i = f_G\left(\mathbf{W}_{(r)}, \frac{\mathbf{L}_{i(r)}}{\mathbf{\Pi}_{(r)}}\right) = 1 + \frac{1}{\sum_r w_r \dfrac{L_{ir}}{\Pi_r}} \sum_{k=1}^{R} w_k \left(w_k \frac{L_{ik}}{\Pi_k} - 2\sum_{j=1}^{k} w_j \frac{L_{ik}}{\Pi_j} \right)$$

$$= \frac{1}{2\sum_r w_r \dfrac{L_{ir}}{\Pi_r}} \sum_{r=1}^{R}\sum_{s=1}^{R} w_r w_s \left| \frac{L_{ir}}{\Pi_r} - \frac{L_{is}}{\Pi_s} \right| \qquad (A.1)$$

Figure A1. Lorenz Curve of Regional Concentration



---

24 Note that the disproportionality measures obtained by decomposing a 0-1 normalized GE measure (see Section 2 and Appendix 3) are not 0-1 normalized.

as the general form of all Gini disproportionality measures of concentration (see Table 1). This general form includes as special cases all Gini coefficients used in the literature on concentration (and specialization). k in equation (A.1) and Figure A1 indexes the observation with the k^{th} lowest region-specific proportionality factor. Note that the expression in the second line of (A.1) is computationally more expensive but does not require sorting observations. The Gini coefficients for the various basic units and references can be defined along the same lines as the *RMD* and *GE* measures discussed in Section 2 (see Table 1).

Appendix 3: Decomposing GE Disproportionality Measures of Localization

Table A1 gives an overview of the measures that can be obtained through decomposition of a Theil index of localization of a nation (country) by state, county, sector and/or industry. The rows in Table A1 determine the level of *regional* aggregation of the measure and the underlying data, which is assumed to be either the nation, the state, or the county level. The columns in Table A1 determine the level of *industrial* aggregation of the measure and the underlying data, which is assumed to be either the total economy, the sector level (agriculture, manufacturing, services), or the industry level. The county-industry level is assumed to be the most disaggregated level for which data is available. The upper left quadrant of Table A1 summarizes all the localization measures, the upper right all the concentration measures and the lower-left all the specialization measures that can be investigated in this setting.

The Theil index given in the upper left cell of the Table, cell (1 [row], 1 [column]), is the Theil index of localization across all counties in the nation and all industries in the economy. It is denoted by $T_{\bullet\bullet}^{loc}(ir)$. The superscript "*loc*" indicates that this measure is a localization measure ("*conc*": concentration; "*spec*": specialization), and the subscript "••" that it covers the entire nation and economy. The term "*ir*" in parentheses indicates that this measure is calculated using data at the level of industries and counties (*j*: sectors; *s*: sectors). $T_{\bullet\bullet}^{loc}(ir)$ is the most comprehensive localization measure. It summarizes all the heterogeneity among the county-industries in a single measure. It can be decomposed in a variety of ways.

To give an example, the overall localization measure, $T_{\bullet\bullet}^{loc}(ir)$, may be stepwise decomposed first by sectors, then by states, and finally by industries. The decomposition by sectors in the first step is informative as regards the contributions to the overall localization of the localizations (across industries and regions) of the individual sectors on the one hand, and the sectoral

Table A1. Measures Obtained Through Decomposition of a Disproportionality Measure of Economic Localization: Theil Index

Regional Level \ Sectoral Level		(1) Measure: total economy; Data: industries	(2) Measure: total economy; Data: sectors	(3) Measure: sector; Data: industries	(4) Measure: total economy; Data: total economy	(5) Measure: sector; Data: sector	(6) measure: industry; data: industry
(1)	Measure: nation; Data: counties	$T^{loc}_{\bullet\bullet}(ir) = \sum_{i=1}^{I}\sum_{r=1}^{R} w_{ir} \frac{X_{ir}}{\overline{X}_{\bullet\bullet}} \ln\left(\frac{X_{ir}}{\overline{X}_{\bullet\bullet}}\right)$	$T^{loc}_{\bullet\bullet}(jr) = \sum_{j=1}^{J}\sum_{r=1}^{R} w_{jr} \frac{\overline{X}_{jr}}{\overline{X}_{\bullet\bullet}} \ln\left(\frac{\overline{X}_{jr}}{\overline{X}_{\bullet\bullet}}\right)$	$T^{loc}_{j\bullet}(ir) = \sum_{i=1}^{I_j}\sum_{r=1}^{R} \frac{w_{ir}}{w_{j\bullet}} \frac{X_{ir}}{\overline{X}_{j\bullet}} \ln\left(\frac{X_{ir}}{\overline{X}_{j\bullet}}\right)$	$T^{conc}_{\bullet\bullet}(\bullet r) = \sum_{r=1}^{R} w_{\bullet r} \frac{\overline{X}_{\bullet r}}{\overline{X}_{\bullet\bullet}} \ln\left(\frac{\overline{X}_{\bullet r}}{\overline{X}_{\bullet\bullet}}\right)$	$T^{conc}_{j\bullet}(\bullet r) = \sum_{r=1}^{R} \frac{w_{jr}}{w_{j\bullet}} \frac{\overline{X}_{jr}}{\overline{X}_{j\bullet}} \ln\left(\frac{\overline{X}_{jr}}{\overline{X}_{j\bullet}}\right)$	$T^{conc}_{i\bullet}(\bullet r) = \sum_{r=1}^{R} \frac{w_{ir}}{w_{i\bullet}} \frac{X_{ir}}{\overline{X}_{i\bullet}} \ln\left(\frac{X_{ir}}{\overline{X}_{i\bullet}}\right)$
(2)	Measure: nation; Data: states	$T^{loc}_{\bullet\bullet}(is) = \sum_{i=1}^{I}\sum_{s=1}^{S} w_{is} \frac{\overline{X}_{is}}{\overline{X}_{\bullet\bullet}} \ln\left(\frac{\overline{X}_{is}}{\overline{X}_{\bullet\bullet}}\right)$	$T^{loc}_{\bullet\bullet}(js) = \sum_{j=1}^{J}\sum_{s=1}^{S} w_{js} \frac{\overline{X}_{js}}{\overline{X}_{\bullet\bullet}} \ln\left(\frac{\overline{X}_{js}}{\overline{X}_{\bullet\bullet}}\right)$	$T^{loc}_{j\bullet}(is) = \sum_{i=1}^{I_j}\sum_{s=1}^{S} \frac{w_{is}}{w_{j\bullet}} \frac{X_{is}}{\overline{X}_{j\bullet}} \ln\left(\frac{X_{is}}{\overline{X}_{j\bullet}}\right)$	$T^{conc}_{\bullet\bullet}(\bullet s) = \sum_{s=1}^{S} w_{\bullet s} \frac{\overline{X}_{\bullet s}}{\overline{X}_{\bullet\bullet}} \ln\left(\frac{\overline{X}_{\bullet s}}{\overline{X}_{\bullet\bullet}}\right)$	$T^{conc}_{j\bullet}(\bullet s) = \sum_{s=1}^{S} \frac{w_{js}}{w_{j\bullet}} \frac{\overline{X}_{js}}{\overline{X}_{j\bullet}} \ln\left(\frac{\overline{X}_{js}}{\overline{X}_{j\bullet}}\right)$	$T^{conc}_{i\bullet}(\bullet s) = \sum_{s=1}^{S} \frac{w_{is}}{w_{i\bullet}} \frac{\overline{X}_{is}}{\overline{X}_{i\bullet}} \ln\left(\frac{\overline{X}_{is}}{\overline{X}_{i\bullet}}\right)$
(3)	Measure: state; Data: counties	$T^{loc}_{\bullet s}(ir) = \sum_{i=1}^{I}\sum_{r=1}^{R_s} \frac{w_{ir}}{w_{\bullet s}} \frac{X_{ir}}{\overline{X}_{\bullet s}} \ln\left(\frac{X_{ir}}{\overline{X}_{\bullet s}}\right)$	$T^{loc}_{\bullet s}(jr) = \sum_{j=1}^{J}\sum_{r=1}^{R_s} \frac{w_{jr}}{w_{\bullet s}} \frac{\overline{X}_{jr}}{\overline{X}_{\bullet s}} \ln\left(\frac{\overline{X}_{jr}}{\overline{X}_{\bullet s}}\right)$	$T^{loc}_{js}(ir) = \sum_{i=1}^{I_j}\sum_{r=1}^{R_s} \frac{w_{ir}}{w_{js}} \frac{X_{ir}}{\overline{X}_{js}} \ln\left(\frac{X_{ir}}{\overline{X}_{js}}\right)$	$T^{conc}_{\bullet s}(\bullet r) = \sum_{r=1}^{R_s} \frac{w_{\bullet r}}{w_{\bullet s}} \frac{\overline{X}_{\bullet r}}{\overline{X}_{\bullet s}} \ln\left(\frac{\overline{X}_{\bullet r}}{\overline{X}_{\bullet s}}\right)$	$T^{conc}_{js}(\bullet r) = \sum_{r=1}^{R_s} \frac{w_{jr}}{w_{js}} \frac{\overline{X}_{jr}}{\overline{X}_{js}} \ln\left(\frac{\overline{X}_{jr}}{\overline{X}_{js}}\right)$	$T^{conc}_{is}(\bullet r) = \sum_{r=1}^{R_s} \frac{w_{ir}}{w_{is}} \frac{X_{ir}}{\overline{X}_{is}} \ln\left(\frac{X_{ir}}{\overline{X}_{is}}\right)$
(4)	Measure: nation; Data: nation	$T^{spec}_{\bullet\bullet}(i\bullet) = \sum_{i=1}^{I} w_{i\bullet} \frac{\overline{X}_{i\bullet}}{\overline{X}_{\bullet\bullet}} \ln\left(\frac{\overline{X}_{i\bullet}}{\overline{X}_{\bullet\bullet}}\right)$	$T^{spec}_{\bullet\bullet}(j\bullet) = \sum_{j=1}^{J} w_{j\bullet} \frac{\overline{X}_{j\bullet}}{\overline{X}_{\bullet\bullet}} \ln\left(\frac{\overline{X}_{j\bullet}}{\overline{X}_{\bullet\bullet}}\right)$	$T^{spec}_{j\bullet}(i\bullet) = \sum_{i=1}^{I_j} \frac{w_{i\bullet}}{w_{j\bullet}} \frac{\overline{X}_{i\bullet}}{\overline{X}_{j\bullet}} \ln\left(\frac{\overline{X}_{i\bullet}}{\overline{X}_{j\bullet}}\right)$	—	—	—
(5)	Measure: state; Data: state	$T^{spec}_{\bullet s}(i\bullet) = \sum_{i=1}^{I} \frac{w_{is}}{w_{\bullet s}} \frac{\overline{X}_{is}}{\overline{X}_{\bullet s}} \ln\left(\frac{\overline{X}_{is}}{\overline{X}_{\bullet s}}\right)$	$T^{spec}_{\bullet s}(j\bullet) = \sum_{j=1}^{J} \frac{w_{js}}{w_{\bullet s}} \frac{\overline{X}_{js}}{\overline{X}_{\bullet s}} \ln\left(\frac{\overline{X}_{js}}{\overline{X}_{\bullet s}}\right)$	$T^{spec}_{js}(i\bullet) = \sum_{i=1}^{I_j} \frac{w_{is}}{w_{js}} \frac{\overline{X}_{is}}{\overline{X}_{js}} \ln\left(\frac{\overline{X}_{is}}{\overline{X}_{js}}\right)$	—	—	—
(6)	Measure: county; Data: county	$T^{spec}_{\bullet r}(i\bullet) = \sum_{i=1}^{I} \frac{w_{ir}}{w_{\bullet r}} \frac{X_{ir}}{\overline{X}_{\bullet r}} \ln\left(\frac{X_{ir}}{\overline{X}_{\bullet r}}\right)$	$T^{spec}_{\bullet r}(j\bullet) = \sum_{j=1}^{J} \frac{w_{jr}}{w_{\bullet r}} \frac{\overline{X}_{jr}}{\overline{X}_{\bullet r}} \ln\left(\frac{\overline{X}_{jr}}{\overline{X}_{\bullet r}}\right)$	$T^{spec}_{jr}(i\bullet) = \sum_{i=1}^{I_j} \frac{w_{ir}}{w_{jr}} \frac{X_{ir}}{\overline{X}_{jr}} \ln\left(\frac{X_{ir}}{\overline{X}_{jr}}\right)$	—	—	—

Notation: $X_{ir} := L_{ir}/\Pi_r$. $T^{type}_{ab}(cd)$ denotes the Theil index of localization (*type*=*loc*), regional concentration (*type*=*conc*), or industrial specialization (*type*=*spec*) for sector or industry *a* (total economy: •) in state or county *b* (whole nation: •), calculated across all industries or sectors *c* in *a* (total economy: •) and/or all counties or states *d* in *b* (whole nation: •). $\overline{X}_{ab} (= \sum_{c \text{ in } a} \sum_{d \text{ in } b} \frac{w_{cd}}{w_{ab}} X_{cd})$ denotes the weighted average of all sectoral units *c* in *a* and all regional units *d* in *b*. X_{cd} may itself be a weighted average across industries and/or counties. w_{cd}, $\Sigma_c\Sigma_d w_{cd} = 1$, denotes the weight of industry or sector *c* in county or state *d*. For example, $\overline{X}_{js} (= \sum_i^{I_j} \sum_r^{R_s} \frac{w_{ir}}{w_{js}} X_{ir})$ in row (5) and column (3) is the weighted average of all industry-county-specific X_{ir} in sector *j* and state *s*; the region-industry-specific weights, w_{ir}, are rescaled by the state-sector-specific weights, $w_{js} = \sum_i^{I_j} \sum_r^{R_s} w_{ir}$, $\Sigma_j\Sigma_s w_{js} = 1$, to standardize the sum of the weights to one. The measures for the empty cells ("—") are zero by definition.

specialization of the country on the other. As a within component, this decomposition yields the weighted average of the Theil indices of localization within each sector, $T_{j\bullet}^{loc}(ir)$, given in cell (1,3) of Table A1. As a between component, it yields the Theil index of sectoral specialization of the nation, $T_{\bullet\bullet}^{spec}(j\bullet)$, given in cell (4,2) of Table A1. In formal terms, this decomposition yields

$$\begin{aligned} T_{\bullet\bullet}^{loc}(ir) &= \sum_{i=1}^{I}\sum_{r=1}^{R} w_{ir}\,\frac{X_{ir}}{\overline{X}_{\bullet\bullet}}\ln\left(\frac{X_{ir}}{\overline{X}_{\bullet\bullet}}\right) \\ &= \sum_{j=1}^{J} w_{j\bullet}\,\frac{\overline{X}_{j\bullet}}{\overline{X}_{\bullet\bullet}}\left[\sum_{i=1}^{I_j}\sum_{r=1}^{R}\frac{w_{ir}}{w_{j\bullet}}\frac{X_{ir}}{\overline{X}_{j\bullet}}\ln\left(\frac{X_{ir}}{\overline{X}_{j\bullet}}\right)\right] + \sum_{j=1}^{J} w_{j\bullet}\,\frac{\overline{X}_{j\bullet}}{\overline{X}_{\bullet\bullet}}\ln\left(\frac{\overline{X}_{j\bullet}}{\overline{X}_{\bullet\bullet}}\right) \qquad \text{(A.2)} \\ &= \sum_{j=1}^{J} w_{j\bullet}\,\frac{\overline{X}_{j\bullet}}{\overline{X}_{\bullet\bullet}}\left[T_{j\bullet}^{loc}(ir)\right] + T_{\bullet\bullet}^{spec}(j\bullet), \end{aligned}$$

where j ($j = 1, \ldots, J$) indexes sectors, I_j denotes the number of industries in sector j, and $w_{j\bullet} = \sum_{i=1}^{I_j}\sum_{r=1}^{R} w_{ir}$ the country-sector-specific weights. Furthermore, $\Sigma_j w_{j\bullet} = 1$; $X_{ir}{:=}L_{ir}/\Pi_{ir}$, $\overline{X} = \sum_{i,r} w_{ir}X_{ir}$, and $\overline{X}_{j\bullet} = \sum_{i=1}^{I_j}\sum_{r=1}^{R} \frac{w_{ir}}{w_{j\bullet}} X_{ir}$.

The decomposition of each sector-specific localization index, $T_{j\bullet}^{loc}(ir)$, by states in the second step is informative as regards the contributions to this sector's localization of the localizations (across industries and regions) of this sector within the individual states on the one hand, and the concentration of this sector across states on the other. As a within component, this decomposition yields the weighted average of the Theil indices of localization by state and sector, $T_{js}^{loc}(ir)$, given in cell (3,3) of Table A1. As a between component, it yields the Theil index of regional concentration of the employment in a sector across states, $T_{j\bullet}^{conc}(\bullet s)$, given in cell (2,5) of Table A1. Formally,

$$\begin{aligned} T_{j\bullet}^{loc}(ir) &= \sum_{s=1}^{S}\frac{w_{js}}{w_{j\bullet}}\frac{\overline{X}_{js}}{\overline{X}_{j\bullet}}\left[\sum_{i=1}^{I_j}\sum_{r=1}^{R_s}\frac{w_{ir}}{w_{js}}\frac{X_{ir}}{\overline{X}_{js}}\ln\left(\frac{X_{ir}}{\overline{X}_{js}}\right)\right] + \sum_{s=1}^{S}\frac{w_{js}}{w_{j\bullet}}\frac{\overline{X}_{js}}{\overline{X}_{j\bullet}}\ln\left(\frac{\overline{X}_{js}}{\overline{X}_{j\bullet}}\right) \\ &= \sum_{s=1}^{S}\frac{w_{js}}{w_{j\bullet}}\frac{\overline{X}_{js}}{\overline{X}_{j\bullet}}\left[T_{js}^{loc}(ir)\right] + T_{j\bullet}^{conc}(\bullet s), \qquad j = 1,\ldots,J, \end{aligned} \qquad \text{(A.3)}$$

where s ($s = 1, \ldots, S$) indexes states, R_s denotes the number of counties in state s, $\overline{X}_{js} = \sum_{i=1}^{I_j}\sum_{r=1}^{R_s}\frac{w_{ir}}{w_{js}}X_{ir}$, $w_{js} = \sum_{i=1}^{I_j}\sum_{r=1}^{R_s} w_{ir}$, and $w_{j\bullet} = \Sigma_s w_{js}$.

Finally, the decompositions of each sector- and state-specific localization measure, $T_{js}^{loc}(ir)$, by industries in the third step is informative as regards the contributions to this state-sector's localization of the concentrations (across regions) of the individual industries within this sector on the one hand, and the industrial specialization of the state-sector on the other. As a within component, this decomposition yields the weighted average of the Theil indices of the regional concentration of each state-industry across counties, $T_{is}^{conc}(\bullet r)$, given in cell (3,6) of Table A1. As a between component, it yields the Theil index of industrial specialization of each state-sector, $T_{js}^{spec}(i\bullet)$, given in cell (5,3) of Table A1. Formally,

$$
\begin{aligned}
T_{js}^{loc}(ir) &= \sum_{i=1}^{I_j} \frac{w_{is}}{w_{js}} \frac{\overline{X}_{is}}{\overline{X}_{js}} \left[\sum_{r=1}^{R_s} \frac{w_{ir}}{w_{is}} \frac{X_{ir}}{\overline{X}_{is}} \ln\left(\frac{X_{ir}}{\overline{X}_{is}} \right) \right] + \sum_{i=1}^{I_j} \frac{w_{is}}{w_{js}} \frac{\overline{X}_{is}}{\overline{X}_{js}} \ln\left(\frac{\overline{X}_{is}}{\overline{X}_{js}} \right) \\
&= \sum_{i=1}^{I_j} \frac{w_{is}}{w_{js}} \frac{\overline{X}_{is}}{\overline{X}_{js}} \left[T_{is}^{conc}(\bullet r) \right] + T_{js}^{spec}(i\bullet), \qquad j = 1,\ldots,J,\ s = 1,\ldots,S,
\end{aligned}
\tag{A.4}
$$

where $w_{is} = \sum_{r=1}^{R_s} w_{ir}$, and $w_{js} = \sum_{i=1}^{I_j} w_{is}$.

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 (2): 231–248.

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

Aiginger, K., and E. Rossi-Hansberg. 2006. Specialization and Concentration: a Note on Theory and Evidence. *Empirica* 33 (4): 255-266.

Allison, P. D. 1978. Measures of Inequality. *American Sociological Review* 43 (6): 865–880.

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: Kluwer.

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

—, and G. Piras. 2008. Spatial Concentration. Unpublished manuscript, University G. d'Annunzio, Chieti, Italy.

Biewen, M. 2002. Bootstrap Inference for Inequality, Mobility and Poverty Measurement. *Journal of Econometrics* 108 (2): 317–342.

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

Bickenbach, F., E. Bode, and C. Krieger-Boden. 2008. Localization in Europe: Stylized Facts. Unpublished manuscript, Kiel Institute for the World Economy, Kiel, Germany.

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. Center of European Integration Studies, Bonn, Germany.

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

—, 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 *Handbook of Urban and Regional Economics*, vol. 4, edited by J. V.Henderson, and J.-F. Thisse. 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.

—. 1989. Sampling Variance and Decomposable Inequality Measures. *Journal of Econometrics* 42 (1): 27–41.

Cowell, F.A. 2000. Measurement of Inequality. In *Handbook of Income Distribution*, vol. 1, edited by A. B. Atkinson, and F. Bourguignon. Amsterdam: Elsevier.

—, and E. Flachaire. 2002. Sensitivity of Inequality Measures to Extreme Values. Discussion paper 60, Distributional Analysis Research Programme, STICERD, London.

Cutrini, E. 2006. The Balassa Index Meets the Dissimilarity Theil Index: A Decomposition Methodology for Location Studies. Quaderno di Ricerca n. 274. Università Politecnica delle Marche, Ancona, Italy.

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

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

—. 2008. Exploring the Detailed Location Patterns of UK Manufacturing Industries using Microgeographic Data. *Journal of Regional Science* 48 (1): 213–243.

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 Specialization and Concentration in the EU. In *Regional Convergence in the European Union*, edited by J. R. Cuadrado-Roura, and M. Parellada. Berlin: Springer.

Iara, A. 2008. The Regional Economic Structure of Romania. In *The Impact of European Integration on Regional Structural Change and Cohesion*, edited by C. Krieger-Boden, E. Morgenroth, and G. Petrakos. London: Routledge.

Krieger-Boden, C. 2008a. Southern Enlargement and Structural Change: Case Studies of Spain, Portugal and France. In *The Impact of European Integration on Regional Structural Change and Cohesion*, edited by C. Krieger-Boden, E. Morgenroth, and G. Petrakos. London: Routledge.

Krieger-Boden, C. 2008b. Eastern Enlargement in a Nutshell: Case Study of Germany. In *The Impact of European Integration on Regional Structural Change and Cohesion*, edited by C. Krieger-Boden, E. Morgenroth, and G. Petrakos. London: Routledge.

Krugman, P. R. 1991. *Geography and Trade*. Leuven: Leuven University Press.

Lafourcade, M., and G. Mion. 2007. Concentration, Agglomeration and the Size of Plants. *Regional Science and Urban Economics* 37 (1): 46–68.

Lasso de la Vega, C., and A. Urrutia. Forthcoming. The Extended Atkinson family: The Class of Multiplicatively Decomposable Inequality Measures, and some new Graphical Procedures for Analysts. *Journal of Economic Inequality*.

Litchfield, J. A. 1999. Inequality Methods and Tools. Unpublished manuscript, London School of Economics, London.

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. Unpublished Manuscript.

Maurel, F., and B. Sédillot. 1999. A Measure of the Geographic Concentration in French Manufacturing Industries. *Regional Science and Urban Economics* 29 (5): 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. European Commission, Brussels.

Mills, J. A., and S. Zandvakili. 1997. Statistical Inference via Bootstrapping for Measures of Inequality. *Journal of Applied Econometrics* 12 (2): 133–150.

Morgenroth, E. 2008. Economic Integration and Structural Change: The Case of Irish Regions. In *The Impact of European Integration on Regional Structural Change and Cohesion*, edited by C. Krieger-Boden, E. Morgenroth, and G. Petrakos. London: Routledge.

Mulligan, G. F., and C. Schmidt. 2005. A Note on Localization and Specialization. *Growth and Change* 36 (4): 565–576.

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 *The Emerging Economic Geography in EU Accession Countries*, edited by I. Traistaru, P. Nijkamp, and L. Resmini. Aldershot: Ashgate.

Openshaw, S., and P. J. Taylor. 1979. A Million or So Correlated Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In *Statistical Applications in the Spatial Sciences*, edited by N. Wrigley, and R. J. Bennet. 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. (ed.). 1999. *Handbook of Income Inequality Measurement*. Boston: Kluwer.

Südekum, J. 2006. Concentration and Specialisation Trends in Germany since Re-Unification. *Regional Studies* 40 (8): 861–873.

Totev, S. 2008. Economic Integration and Structural Change: The Case of Bulgarian Regions. In *The Impact of European Integration on Regional Structural Change and Cohesion*, edited by C. Krieger-Boden, E. Morgenroth, and G. Petrakos. London: Routledge.

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