IIIS Discussion Paper

No.2 / July 2003

An Account of Geographic Concentration Patterns in Europe

Marius Brülhart
DEEP-HEC, University of Lausanne

Rolf Traeger
United Nations Economic Commission for Europe, Geneva

IIIS Discussion Paper No. 02

An Account of Geographic Concentration Patterns in Europe

Marius Brülhart

Rolf Traeger

An Account of Geographic Concentration Patterns in Europe*

Marius Brülhart†
DEEP-HEC, University of Lausanne

Rolf Traeger‡
United Nations Economic Commission for Europe, Geneva

This version: July 14, 2003

Abstract

This paper provides a methodologically rigorous description of sectoral location patterns across Western European regions over the 1975-2000 period. To measure geographic concentration, we use decomposable entropy indices and associated bootstrap tests. In addition, we estimate locational centre-periphery gradients for individual sectors and the impact of EU membership on countries' internal geography. It is found that manufacturing has become gradually and statistically significantly more concentrated, although the locational bias towards central regions has become weaker. Conversely, market services have been relocating towards centrally located regions. Accession to the EU has strengthened countries' internal concentration trends.

JEL Classification: R12, R14, F15

Keywords: geographic concentration, EU regions, centre-periphery gradients, entropy indices, bootstrap inference

<All tables and figures at end>

*We would like to thank Philip Lane for helpful comments and Martin Hallet for the generous provision of data. Financial support from the Swiss National Science Foundation (grant 612-65970) and from the UK Economic and Social Research Council (ESRC grant L138251002) is gratefully acknowledged.

†Correspondence address: Département d'économétrie et économie politique, Ecole des HEC, Université de Lausanne, CH - 1015 Lausanne, Switzerland; Marius.Brulhart@hec.unil.ch; http://www.hec.unil.ch/mbrulhar/

‡Any opinions expressed in this paper are those of the authors and do not necessarily reflect those of UNECE or its member countries.

1 Introduction

The spatial analysis of integrating market economies has recently regained prominence on the economic research agenda. This has two main reasons. First, as policy initiatives and technological advances have conspired over the last half-century to reduce the costs of economic transactions across region and country borders, economic activities are generally believed to have become increasingly "footloose". Second, theorists have made substantial progress in the 1990s in modelling location forces that are not due to underlying spatial heterogeneity but to the interplay between market forces and distance costs in homogeneous space. The "new economic geography" provides a formal treatment of agglomeration and dispersion forces in such a world.[1] One of the most interesting insights of this literature is that economic integration may render some activities less rather than more "footloose", because falling trade costs can contribute to a strengthening of agglomeration economies.

Both the policy-related and the theory-based motivations for renewed interest in spatial economics are particularly relevant to Western Europe, which has gone through a process of unprecedented economic integration, and where underlying endowment differences are small compared to more resource-dependent world regions. Considerable research effort has therefore been expended on studying location patterns of sectoral production and employment in Europe.[2] It has proven difficult to distil strong stylised facts from this research. One reason for the heterogeneity of results is that the studies differ quite strongly in the data and measures they employ. More fundamentally, it appears that sectoral relocation in Europe is a slow and multifaceted process that does not leap out from the data. Overman, Redding and Venables (2001) have summarised the dominant view as follows: "In contrast to the US, EU countries are becoming increasingly specialised (...), although the changes are not particularly large." This diagnosed tendency towards increased specialisation applies to the distribution of manufacturing sectors across countries - little is still known about geographic concentration of sectors at sub-national level and across the full range of economic activities.

The aim of this paper is to provide a comprehensive and methodologically rigorous account of sectoral concentration patterns across Western European regions, in a quest for empirically well-founded stylised facts. Our study distinguishes itself from the existing literature in five principal respects.

First, we apply entropy indices to measure geographic concentration. These in-

[1] See Fujita, Krugman and Venables (1999) and Fujita and Thisse (2002) for comprehensive statements.

[2] For studies of geographic concentration patterns in Europe using sectoral output or employment data, see Aiginger and Leitner (2002), Aiginger and Pfaffermayr (2003), Amiti (1999); Brülhart (2001a, 2001b); Clark and van Wincoop (2001); Haaland, Kind, Midelfart Knarvik and Torstensson (1999); Hallet (2000); Helg, Manasse, Monacelli and Rovelli (1995); Imbs and Wacziarg (2003); Kalemli-Ozcan, Sorensen and Yosha (2003); Krugman (1991); Midelfart Knarvik, Overman, Redding and Venables (2000); Peri (1998); and Storper, Chen and De Paolis (2002).

dices have distinct advantages over the conventional measures in this literature. One of the advantages lies in their suitability to inequality decomposition analysis. This allows us to compare within-country concentration to between-country concentration in conceptually rigorous fashion. In addition, we can quantify how much each sector contributes to aggregate geographic concentration, by decomposing aggregate entropy into the "factor contributions" of individual sectors.[3]

Second, we employ bootstrap inference to test the statistical significance of changes in observed concentration measures. These tests have been shown to be particularly accurate when used in conjunction with entropy measures.

Third, we address aggregation biases that arise in regional data and are often overlooked. Consideration of this issue leads us to compute separate indices for "relative concentration", where we measure the degree to which sectors are concentrated relative to the geographic distribution of aggregate activity, and for "topographic concentration", where we measure the degree to which sectors are concentrated relative to physical space. Our results show that this conceptual distinction has substantial empirical relevance.

Fourth, we use regression techniques to estimate (a) the degree to which sectoral location patterns are influenced by the centrality and peripherality of regions and (b) whether and to what extent accession to the EU has affected the time profiles of within-country location patterns.

Fifth, our study is based on comprehensive regionally and sectorally disaggregated data sets. Our main data set provides us with a balanced panel of employment in eight economic sectors in 236 NUTS-2 and NUTS-3 regions belonging to 17 Western European countries over the 1975-2000 period.[4] The eight sectors of this data set cover the full range of economic activities, including agriculture and services. Through the use of employment as the size measure we can avoid problems of currency conversion inherent in value data. As a complement to the main data set, we use a data set that disaggregates manufacturing value added into nine industries for 116 EU-15 NUTS-1 and NUTS-2 regions over the 1980-1995 period.

We have several motivations for studying the relative magnitude of intra- and international specialisation trends. One motivation stems from the fact that this distinction has considerable policy relevance. For example, the desirability for a country to adopt the single currency hinges on the degree of country specificity of economic shocks. To the extent that shocks are sector specific, inter-country specialisation will increase the asymmetry of shocks and thereby reduce the attractiveness

[3] An entropy index has previously been used to measure geographic concentration by Aiginger and Davies (2000), who applied the index to country-level output data for the EU. They did not make use of the index's decomposability. A number of studies have also applied entropy measures and their decompositions to describe the spatial inequality of aggregate income (see e.g. Duro and Esteban, 1998).

[4] NUTS (Nomenclature of Territorial Units for Statistics) is Eurostat's classfication of sub-national spatial units, where NUTS-0 corresponds to the country level and increasing numbers indicate increasing levels of sub-national disaggregation.

of monetary union. If specialisation were mainly an intra-country phenomenon, however, it would be of no consequence for the cost of monetary union. Second, there is an ongoing debate about the extent to which regional policy should fall in the remit of national governments or in that of supranational European Union authorities. To the extent that regional policy targets certain sectors or that specialisation patterns affect relative income levels, inter-country specialisation will strengthen the case for delegating regional policy to a supranational authority, while intra-country specialisation is arguably better addressed by national policy makers.[5]

Our paper is organised as follows. Section 2 presents the measures used, their associated bootstrap tests, and our data resources. In Section 3, we describe geographic concentration patterns using the entropy measures, and in Section 4 we apply regression techniques to estimate locational centre-periphery gradients and the impact of EU accession. Section 5 provides a concluding summary and discussion.

2 Measurement, inference and data

Following Krugman (1991), "locational Gini indices" have become the measure of choice for studies of geographic specialisation patterns. The Gini index has strong intuitive and pedagogical appeal, but it is not ideally suited to our analysis. One feature that we seek in a measure of geographic concentration is decomposability into its within-country and between-country components. The Gini index is only decomposable if the range of the values taken by the variable of interest does not overlap across subgroups of individual observations (Cowell, 1980). This is evidently not the case in our context: regions in different countries may well have similar degrees of specialisation in a particular sector. Another desirable characteristic of any retained measure would be its suitability for statistical inference.

It turns out that measures that pertain to the single-parameter generalised entropy class (GE(α)) perform particularly well on both those counts. Entropy measures have the welcome feature of being additively decomposable both by population subgroup and by factor components. In addition, entropy measures lend themselves particularly well to bootstrap-based statistical inference.[6]

[5] Giannetti (2002) has found that the sectoral composition of EU regions affects those regions' growth trajectories and thus helps explain the coexistence of inter-country income convergence and intra-country income divergence. Also note that one criterion for "objective 2" status and the associated eligibility for regional aid from the EU is a higher percentage of jobs in industry than the EU average and a decline in industrial employment.

[6] We have computed Gini indices as well as entropy measures, where applicable. The choice of index did not affect our qualitative findings, and we therefore report only the entropy-based results. The results are available from the authors on request.

2.1 Additively decomposable inequality measures: General entropy

Before giving a technical presentation of the relevant entropy measures, we should alert readers that we judge it most pedagogical to introduce those measures first in an abstract sense, and only then to flesh them out with our concrete empirical context. Hence, this subsection presents measurement issues without reference to our geographic applications. The geographic context will be introduced in the following subsection. All concepts presented here will later be taken up in the empirical analysis.

In abstract terms, the underlying concepts are as follows. We consider a population of *basic units* $i \in \{1, 2, ..., N\}$, where each basic unit is associated with a unique value of the measured variable y, and $\sum_{i=1}^{N} y_i \equiv Y$. Then, we define an exhaustive partition of this population into mutually exclusive *subgroups* of basic units $k \in \{1, 2, ..., K\}$. Moreover, the variable y is defined such that it can be subdivided exhaustively into mutually exclusive *factors* $f \in \{1, 2, ..., F\}$.[7]

Members of the generalised entropy (GE) class of inequality indices are defined by the following expression:

$$GE(\alpha) = \frac{1}{\alpha^2 - \alpha} \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_i}{\overline{y}} \right)^{\alpha} - 1 \right] \tag{1}$$

where

$$\overline{y} = \frac{1}{N} \sum_{i=1}^{N} y_i = \frac{Y}{N},$$

and α is a sensitivity parameter. α measures the weight given to distances among values taken by y at different parts of the distribution of y. It can in principle be set to any real number. The neutral parameter value is 1. If $\alpha < 1$, then a bigger weight is attributed to the dispersion of y in the lower tail of the distribution of y over i, and if $\alpha > 1$, then a bigger weight is attributed to the dispersion in the upper tail. Like the Gini, these indices increase in the degree of inequality.

Following standard practice, we confine our analysis to the cases where $\alpha = 1$ and $\alpha = 2$. Using L'Hopital's rule on equation (1), the first case yields the *Theil index* of inequality:

$$\mathrm{GE}(1) = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i}{\overline{y}} \log \frac{y_i}{\overline{y}}, \tag{2}$$

where

$$0 \leq GE(1) \leq \log N.$$

[7] In the income distribution literature, where these measures were first used by economists, i would typically refer to individuals, y to income, k to socio-economic categories and f to different income sources (wages, government transfers, capital income, etc.). The definitions of these concepts in the context of our study will be provided in the next subsection.

The second case yields *half the squared coefficient of variation, CV*:

$$\text{GE}(2) = \frac{1}{2} CV^2, \tag{3}$$

where

$$CV = \frac{1}{\overline{y}} \left[\frac{1}{N} \sum_{i=1}^{N} (y_i - \overline{y})^2 \right]^{\frac{1}{2}},$$

and

$$0 \leq GE(2) \leq \frac{1}{2}(N-1).$$

A simple illustration of the behaviour of these measures is given in Appendix 1.

These indices are **decomposable by population subgroups** in particularly appealing fashion. Each GE index can be decomposed additively as:

$$GE(\alpha) = GE_w(\alpha) + GE_b(\alpha), \tag{4}$$

where GE_w and GE_b stand for within-subgroups and between-subgroups general entropy respectively.

Between-group inequality, GE_b, is computed by applying equation (1) to the K subgroup means $\overline{y}_k$ instead of the N observations on y.

The contribution of within-subgroup inequality is computed as follows:

$$GE_w(\alpha) = \sum_{k=1}^{K} \left(\frac{N_k}{N} \right)^{1-\alpha} \left(\frac{Y_k}{Y} \right)^{\alpha} GE_k(\alpha), \tag{5}$$

where $GE_k(\alpha)$ is the GE index as defined by equation (1) but confined to observations belonging to subgroup k (so that N becomes N_k). Subgroup GE indices are therefore calculated as if each subgroup were a separate population.

It is evident from equation (5) that the GE(1) index weights subgroup inequalities by the y shares. The GE(2) index decomposition implies weights that are based on the n shares as well as the y shares. For decompositions by population subgroups, GE(1) is generally preferred to GE(2), because for GE(2) the weights used to compute GE_w are not independent from GE_b.[8]

For a **decomposition of overall inequality by factors**, we seek a rule according to which we can express a measure of total inequality in y, which we denote I, as the sum of the contributions from all factors, so that factor f provides a disequalising contribution if $S_f > 0$, and an equalising contribution if $S_f < 0$:

[8] Bourguignon (1979) and Shorrocks (1980) have proven that GE(0) and GE(1) are the only additively decomposable scale invariant inequality measures for which the weigths of the within-subgroup inequalities sum to a constant (i.e. 1) and are independent of GE_b. Shorrocks (1984) showed that even if one relaxes the *additively* decomposable constraint by allowing weaker aggregation properties, the admissible set of indices expands only to monotonic transformations of the $GE(\alpha)$ family.

$$I = \sum_{f=1}^{F} S_f(I).$$

Functions that generate suitable values of factor contributions S_f are referred to as "decomposition rules". The adoption of such a rule is necessary to apportion inequality contributions exhaustively and uniquely to individual factors when the y-contributions from different factors are correlated. In general, there is an infinite possible number of such rules, and the choice is arbitrary. However, Shorrocks (1982) has proven that under some weak and plausible assumptions one arrives at the following unique decomposition rule for proportional factor contributions $s_f(I)$:

$$s_f(I) = \frac{S_f(I)}{I} = \rho_f * \frac{\sigma(\mathbf{y}_f)}{\sigma(\mathbf{y})} = \frac{\mathrm{cov}(\mathbf{y}_f, \mathbf{y})}{\sigma^2(\mathbf{y})},$$

where $\mathbf{y} = (y_1, ..., y_N)$ is the vector of total y's, $\mathbf{y}_f = (y_{f1}, ..., y_{fN})$ is the vector of y's from factor f, σ is the standard deviation, and ρ_f is the correlation between $\mathbf{y}_f$ and $\mathbf{y}$.[9] This decomposition rule is especially appealing, since, as shown by Shorrocks (1982), it yields the same set of proportional factor contributions s_f irrespective of the inequality index I that is chosen. In terms of the proportional factor contributions, the choice of inequality measure therefore becomes irrelevant. However, it is standard practice to resort in this context to the GE(2) index, for which the Shorrocks decomposition rule happens to be the "natural rule", since:

$$s_f = \rho_f \frac{\overline{y_f}}{\overline{y}} \sqrt{\frac{\mathrm{GE}(2)_f}{\mathrm{GE}(2)}}. \tag{6}$$

Hence, a certain factor f's proportional contribution to total inequality is the product of (a) the correlation of $\mathbf{y}_f$ with $\mathbf{y}$, (b) f's share in total y, and (c) the inequality in that factor relative to total inequality, measured using GE(2).[10]

2.2 The spatial aggregation problem: Topographic versus relative concentration

We now relate the concepts introduced in the previous subsection to our specific empirical context. Most of those concepts have a straightforward geographic counterpart: for "inequality" read "geographic concentration"; y represents economic activity (measured e.g. by employment or value added); the subgroup variable k denotes countries; and the factor variable f denotes economic sectors. Decomposition by population subgroups therefore corresponds to an analysis of intra-country

[9] s_f, of course, corresponds to the slope coefficient from a regression of $\mathbf{y}_f$ on $\mathbf{y}$.

[10] S_f can be interpreted in two different ways; (a) as the inequality that would be observed if factor f were the only source of inequality in y, S_f^A, and (b) as the amount by which total inequality would change if inequality in terms of factor f were reduced to zero, S_f^B. Shorrocks (1982) has shown that, for the GE(2) index, $S_f = \frac{1}{2}(S_f^A + S_f^B)$, whereas for most other inequality indices there exists no such obvious connection between S_f and (S_f^A, S_f^B).

and inter-country geographic concentration, and decomposition by factors quantifies to what extent each sector contributes to the geographic concentration of total economic activity.

The one concept that is difficult to define in the geographic context is that of a "basic unit". In studies of income inequality, this issue is uncontroversial: each person is a basic unit. Applied to geographic analysis, the definition of a basic unit is less obvious. Our most disaggregated observed units are NUTS-2 and NUTS-3 regions, i.e. sub-national spatial units of European countries (see Appendix 2). These regions, the "observed units" in our data, should not be interpreted as the basic units, because they differ significantly in terms of both geographic and economic size, and it is well known that spatial inequality measures are sensitive to the definition of regions. This is commonly referred to as the "modifiable areal unit problem" (MAUP), according to which the results of statistical analysis of data for spatial zones can be varied at will by changing the zonal boundaries (Arbia, 1989). The problem has two components; a problem of scale, involving the aggregation of smaller units into larger ones, and a problem of alternative allocations of component spatial units to zones (gerrymandering).

To acknowledge that regions should not be used as basic units still leaves open a large number of possible definitions. We use two definitions of a basic unit: a square kilometre, and a unit of economic activity (which can mean either an employed person or a unit of value added). The choice of definition may seem innocuous, but in fact it implies fundamentally different underlying meanings of "geographic concentration". Our results show that empirical results are highly sensitive to this choice.

When we define a basic unit as a square kilometre (or any other areal unit), the no-concentration benchmark obtains where an activity is spread perfectly evenly across geographic space. Conversely, any departure from such an even spatial spread will register as geographic concentration, irrespective of the spatial distribution of endowments or of other economic sectors. We refer to this conception of geographic concentration as "*topographic concentration*".[11]

If we use the alternative definition of a basic unit as "an employed person", then we condition topographic space by the distribution of overall employment. In this case, the no-concentration benchmark implies that each (co-located group of X) employed person(s) allocates her (their) working time across sectors exactly according to the proportions corresponding to those sectors' use of employed labour across all locations. This is the concept of concentration that has been used in most

[11] Note that this definition differs from the concept of "absolute" concentration, where basic units are defined as corresponding exactly to the observed spatial units, i.e. regions or countries (Aiginger and Leitner, 2002; Aiginger and Pfaffermayr, 2003; Haaland *et al.*, 1999). As pointed out by Combes and Overman (2003), the no-concentration benchmark implied by "absolute" concentration is that an industry has identical employment/output in all regions irrespective of those regions' size, which is difficult to reconcile with any market-based location model.

previous studies and that seems economically most relevant. We shall refer to this definition as "*relative concentration*". Hence, given the spatially uneven distribution of aggregate employment, a sector that happens to be perfectly evenly spread in space would have zero topographic concentration but positive relative concentration. Conversely, a sector that is spread exactly proportionally to total activity would have zero relative concentration but positive topographic concentration. Where we calculate relative concentration using value added rather than employment data, we condition space on the distribution of overall value added. In this case, a basic unit corresponds to one unit of value added, irrespective of the sector that generates that unit. Note, finally, that we use the expression "geographic concentration" as the general term that encompasses both the "topographic" and the "relative" definition.

Formally, our observed regions $r \in \{1, 2, ..., R\}$, are sets of basic units i, and we refer to them as *observed units*.[12] The size of each observed unit is defined in terms of the number of basic units it contains, $n_r \geq 1$, such that $\sum_r n_r = N$. The observed variable Y_r corresponds to observed-unit totals of unobserved basic-unit realisations of y ($Y_r = \sum_i y_{ir}$). With countries as our subgroups k, we can write that $N > R > K$.

In this setting, the expressions for the two basic entropy indices become:

$$\mathrm{GE}(1) = \sum_{r=1}^{R} \frac{n_r}{N} \frac{\overline{y}_r}{\overline{y}} \log \frac{\overline{y}_r}{\overline{y}} = \sum_{r=1}^{R} \frac{Y_r}{Y} \log \frac{\overline{y}_r}{\overline{y}}, \tag{7}$$

and:

$$CV = \frac{1}{\overline{y}} \left[\sum_{r=1}^{R} \frac{n_r}{N} \left(\overline{y}_r - \overline{y}\right)^2 \right]^{\frac{1}{2}}, \tag{8}$$

where

$$\overline{y}_r = \frac{Y_r}{n_r}, \text{ and } \overline{y} = \frac{Y}{N},$$

and where n_r, depending on whether we measure topographic or relative concentration, corresponds to regions' land area or total economic activity (in either employment or value added terms). Simple illustrations of the behaviour of these measures for both changes in Y_r and changes in n_r are given in Appendix 1.

These measures are true representations of actual inequality only if inequality among basic units inside observed units is zero. If intra-regional inequality exists, which of course applies in reality, the weighted measures will underestimate total inequality. This downward bias in measured inequality rises with the level of spatial aggregation. It is a manifestation of the scale-related MAUP. By size-weighting the GE indices in expressions (7) and (8), we minimise the downward bias given the

[12] In the income distribution context, r could for instance correspond to households.

data at hand, but we cannot eliminate it.[13]

For the second component of the MAUP, the arbitrariness inherent in administrative region borders, given a certain distribution of region sizes, there is no methodological palliative. In addition, broad statistical definitions of sectors may also obscure economically relevant concentration patterns, if offsetting concentration structures of sub-sectors are blurred by the aggregation of those sub-sectors. Absolute levels of the indices, and decompositions thereof, must therefore be interpreted with caution. However, the focus of this study is on changes in geographic concentration patterns over time, and if biases due to the MAUP and to sectoral aggregation are stable intertemporally, their absolute magnitude will not distort our inference.[14]

2.3 A bootstrap test for the significance of changes in geographic concentration

Any concentration index describes the dispersion of a distribution through a scalar, and it therefore has its own sampling distribution. Traditionally, inference on inequality measures has been based on asymptotic results obtained through the delta method. For a test of the equality of two distributions on the same units at different times, however, this method requires cumbersome covariance calculations to take account of the intertemporal dependencies in the data. Furthermore, the finite-sample properties of such tests are unknown.

Hence, Biewen (2001) and Mills and Zandvakili (1997) have argued in favour of using bootstrap inference. With this approach, the sampling distribution of an inequality index is estimated by multiple random resampling with replacement from the data set at hand. Through the bootstrap one can account for dependencies in the data without having to estimate covariance matrices explicitly. Biewen (2001) proved that the bootstrap test for inequality changes over time is consistent for any inequality statistic that can be expressed in terms of population moments - which includes the GE class of indices but not the Gini index. This result is shown by

[13] One approach used in the income inequality literature to deal with grouped data is to estimate a certain distribution function parametrically using maximum likelihood, and to calculte inequality indices over the estimated distribution. We do not follow this route for two reasons. First, we have no priors as to the functional form of such a distribution. Second, there is no clear case based on empirical work for favouring either our non-parmetric approach or the parametric method (Slottje, 1990).

[14] In the income inequality literature, there is evidence that ignorance about intra-household inequality biases inequality measures downwards significantly, but that these biases have negligible impact on cross-sectional comparisons (Haddad and Kanbur, 1990). However, evidence on the co-location of firms at the micro-geographic level points to the importance of narrowly confined clusters. According to Duranton and Overman (2002), the relevant distance for geographical clusters of British manufacturing firms is mostly smaller than 50 kilometers. In comparison, the radius of a circle with a surface corresponding to the average area of regions in our data set 1 (15,000 km^2) is 69 kilometres. The degree of accuracy with which regional data reflect patterns and changes in these fundamental distributions remains to be studied systematically.

Biewen to be valid also for grouped data (i.e. for observed units that are aggregates of basic units). Using Monte Carlo simulations, he demonstrated that this approach achieves a finite-sample coverage accuracy that is equivalent to that obtained through analytically derived (but asymptotic) tests. Mills and Zandvakili (1997) found that the bootstrap estimated standard errors were closer to the corresponding asymptotic estimates for the Theil index than for the Gini index, and they too therefore preferred the entropy measure.

The standard use of the bootstrap is as a method for making probabilistic statements about population parameters based on a data sample drawn randomly from that population. One interpretation of this test in our context is therefore to consider our yearly sets of regional observations as random draws from the universe of (industrialised) world regions. Alternatively, one could consider the set of Western European regions as the population, and search for specifically Western European parameters. In this setting, bootstrap inference remains useful, considering that the data are measured with error, and that the measurement error is distributed stochastically across observations (assuming that measurement errors are distributed independently from y). The principal attraction of the bootstrap in this case is that it absolves us from making assumptions on the form of the measurement error distribution across observations.[15]

By treating all observations equally in the resampling process, the standard bootstrap method implies that the measurement errors attached with each observation are *iid* draws from the population error distribution. This assumption is difficult to justify in the context of our study, as we have strong reason to believe that measurement errors are to a large extent country-specific (i.e. spatially autocorrelated). We therefore apply block-wise resampling, defining countries as blocks. For each replication, a sample is drawn randomly among K blocks of regions, where each block has sample size R_k. Since we have no priors on the distribution of measurement errors across countries, we attach equal probability weights to those K sets of observations in the resampling procedure.[16] All bootstrap results are based on 10,000 replications.

2.4 Data

We draw on two complementary data sets, both of which are described in detail in Appendix 2. Data set 1, compiled by Cambridge Econometrics, provides a balanced

[15] An alternative strategy for inference on concentration indices in exhaustive samples of grouped data with measurement error is to assume certain distributions of those measurement errors and to simulate corresponding distributions for the concentration indices (Bourguignon and Morrison, 2002). That approach requires strong assumptions on the distributional forms of measurement errors.

[16] We ran all tests also with region-level resampling. As expected, this yielded generally tighter confidence intervals, but the higher moments of the distributions underlying those intervals were not affected significantly.

panel of sectoral employment for 17 West European countries, the 15 EU member states plus Norway and Switzerland (collectively referred to as WE17). Except for Luxembourg, all country data are disaggregated into NUTS-2 or NUTS-3 regions, giving a total of 236 region-level observations per sector and year. The number of regions within countries ranges from 2 (Ireland) to 37 (UK). Employment is reported for eight sectors, covering the full range of economic activities, over the period 1975-2000.

Figure 1 illustrates the evolution over our sample period of the relative sizes of the eight sectors in data set 1. It emerges clearly that the WE17 economies have been marked in the last quarter century by pronounced growth in the relative sizes of the tertiary sector, at the expense of the primary and the secondary sectors. This fact alone provides strong motivation for studying geographic specialisation patterns not just for manufacturing industries, but across the full spectrum of economic activities.

Data set 2, compiled by Hallet (2000), reports gross value-added (GVA) of nine manufacturing sectors across the 15 EU member states (referred to as EU15). For eight countries, the data are disaggregated into either NUTS-1 or NUTS-2 regions, giving a total of 109 regions. The remaining seven countries appear in the data as single regions. Among the countries that are subdivided, the number of regions ranges from 5 (Portugal) to 23 (UK). The period covered is 1980-1995.

The two data sets differ in terms of geographic and sectoral disaggregation, but they are complementary. The time span of the second is encompassed by that of the first. Moreover, data set 1 offers a broader base for comparison of agglomeration between and within countries, because it is more regionally disaggregated. We consider employment data as preferable to data based on production values, because the former are not subject to the problems associated with price conversions across countries and years. The comparative attraction of data set 2 is the detail it provides on manufacturing sectors, which facilitates comparisons with previous research findings by bringing us closer to the data sets that have been used in most existing studies.

We complement those data sets with a vector of "peripherality indices" for our sample regions, as computed by Copus (1999). These indices range from 0 (most central region) to 100 (most peripheral region) and are derived from inversely distance-weighted averages of regional GDPs.[17] The underlying interregional distances were quantified on the basis of a regional matrix of road-freight travel times, and GDPs are measured in a common currency using purchasing-power parity exchange rates.

[17] See equation (9) below. The regional breakdown used by Copus (1999) is in most cases finer than that of our study. Hence, we aggregated up peripherality indices of sub-regions using GDP weights. In our data set, the region with the lowest peripherality index is Inner London (21), and the one with the highest index is Northern Norway (100).

3 Geographic concentration: Regions versus countries

3.1 Relative concentration across all regions

3.1.1 All sectors

Sectoral Theil indices of relative concentration across the full spectrum of activities in WE17 regions (i.e. using data set 1) are reported in Table 1 and Figure 2. These indices are computed according to equation (7) using total regional employment as the weighting variable n_r.

On average over our sample period, agriculture turns out to be by far the most concentrated sector (note the log scale of Figure 2), and manufacturing is second-most concentrated, while construction is the most dispersed.

These results seem plausible. In view of the regional and sectoral aggregation problems, however, our analysis focuses not on levels but on changes over time. In Table 1, we report changes in relative concentration (i) over our entire sample period 1975-2000, (ii) over the subperiod 1975-1987 and (iii) over the subperiod 1987-2000. The sample period is divided in this way since 1987 coincides with the entry into force of the Single European Act and thus the launch of the EU's Single Market programme. Hence, one can interpret the second subperiod as a time of particularly strong policy-led integration. Table 1 also reports statistical significance levels according to the bootstrap test described above.

We find that manufacturing is the only sector that has seen a monotonic and statistically significant increase in relative concentration. This increase was more pronounced in the post-Single Market subperiod than in the earlier subperiod. Our analysis therefore confirms the general finding of the previous literature that European manufacturing is becoming more geographically concentrated, particularly since the inception of the Single Market programme.[18]

Our results of Table 1 furthermore show that, with the exception of the "transport and communications" industry, which has become significantly more dispersed, no service sector exhibits a statistically significant change in relative concentration over the full sample period. On the whole, therefore, the evidence does not support the view of strong sectoral reallocation trends across the spectrum of economic activities. Looking at the subperiods, however, we find that the tendency to concentrate (disperse) geographically is stronger (weaker) in the second subperiod than in the first subperiod for all eight sectors. This finding is consistent with the view that the deepening of European integration through the Single Market programme has favoured geographic concentration forces.

[18] We estimate the association between EU membership and geographic concentration trends explicitly in Section 4.

3.1.2 Manufacturing

In Table 2, we report indices of relative concentration for disaggregated manufacturing sectors across EU15 regions, calculated from our data set 2. As noted above, these findings are not strictly comparable with those based on data set 1, due to differences of measurement units (value added instead of employment) and to narrower regional and time coverage.

The results of the two data sets are consistent in so far as they both show a trend towards stronger relative concentration of total manufacturing for the first subperiod (although not for the second one). The strongest increase in relative concentration is found for the textiles, clothing and footwear sector - a tendency which is particularly pronounced in the post-1987 subperiod but statistically significant throughout. This confirms earlier findings whereby the strongest relocation tendencies in European manufacturing are in relatively low-tech and labour-intensive sectors. We do not find a statistically significant change in the concentration index over the full 1980-1995 period for any other manufacturing sector. Six of the nine sectors display stronger concentration trends post-1987 than pre-1987. Here too, we can therefore retain as a stylised fact that EU industries exhibit weak overall concentration pressures, with some evidence of a strengthening subsequent to 1987.[19]

3.2 Relative concentration: Between-country and within-country components

Exploiting the decomposability of entropy indices according to equation (4), we can track the evolution of the within-country and between-country components of geographic concentration.[20]

3.2.1 All sectors

Using data set 1, we have computed within-country shares of relative concentration ($GE_w(1)/GE(1)$) across all sectors. The results are reported in Figure 3.

On average, most of the concentration of service sectors is between countries rather than within countries. The opposite applies to manufacturing: within-country concentration largely dominates between-country concentration.

[19] According to the last row of Table 2, total manufacturing seems to have become more concentrated pre-1987 and more dispersed thereafter, which is not consistent with the concentration time profile found in the employment data. However, this is driven largely by "machinery, electrical and electronics", the largest of our nine manufacturing industries, for which we find a significant initial increase and a significant subsequent decrease in concentration. Inspection of the data suggests the post-1987 decrease is primarily driven by a drop in reported value added of this sector in the West German regions. Given the estimated nature of the statistics for Germany in our data set 2, this result might be influenced by measurement problems (see Hallet, 2000).

[20] In the context of relative concentration, a "factor decomposition" of total concentration is meaningless, since the concentration of total employment across regions weighted by total employment is zero.

In terms of changes over time, the within-country share of relative concentration has fallen over our sample period for a majority of sectors. Hence, between-country concentration forces seem to have been relatively stronger than within-country concentration forces. Given that countries' internal markets were already liberalised in 1975, whereas our sample period was marked by strong between-country liberalisation, this result is in line with the view that European integration opens scope for between-country specialisation which hitherto had existed only at the within-country level.

Relative concentration of manufacturing exhibits a trend break in the early 1990s towards a re-increase in the within-country share. It thus appears that, after a period of more pronounced inter-country concentration processes, intra-country agglomeration forces have come to dominate relocation of manufacturing employment in the 1990s.

3.2.2 Manufacturing

Within-country shares of relative concentration for the manufacturing sectors, based on data set 2, are given in Figure 4. In this data set too, the within-share of relative concentration of total manufacturing shows a u-shaped time profile - declining in the 1980s but increasing since the early 1990s.

The industry that emerges with the clearest trend is textiles, clothing and footwear, which exhibits a steady decline in the within-country share of geographic concentration.

3.3 Topographic concentration across all regions

The choice of spatial weights, which might seem an arcane technicality, turns out to be empirically important. Table 3 and Figure 5 report indices of topographic concentration, computed for data set 1. The difference compared to the relative concentration indices is most evident for agriculture. Of our eight sample sectors, agriculture exhibits the highest average level of relative concentration but the lowest level of topographic concentration. In both cases the gap separating agriculture from the most similarly concentrated sector is large. These results are of course entirely consistent. While agriculture is spread out more than the other sectors in line with total land area, it is typically concentrated in regions with low employment densities, and hence it is concentrated strongly when we condition the spatial distribution of agricultural employment by the distribution of total employment. Another difference between topographic and relative concentration is that service sectors are by far the most concentrated ones in the former case, whereas in terms of relative concentration they are less concentrated than manufacturing as well as agriculture.

Turning to the time profiles of our topographic concentration measures, Figure 5 suggests that the topographic concentration of total employment has remained stable over the sample period, and the bootstrap test does not reject the null hypothesis of identical concentration indices in the base and end periods.

The evident stability in the topographic distribution of total employment, however, masks offsetting changes in the topographic concentration of individual sectors. The most pronounced trends are an increase in topographic concentration of agriculture and a simultaneous decrease in the concentration of manufacturing. These changes are statistically significant. The decrease in topographic concentration of manufacturing, together with the detected increase in relative concentration, suggests that manufacturing has relocated from regions with high employment density to regions with low employment density.

3.4 Topographic concentration: Decompositions

3.4.1 Between-country and within-country components

The decomposition of aggregate topographic concentration into its within-country and between-country components is reported in Figure 6. On average, service sectors have the highest share of within-country concentration, again as opposed to the patterns observed for relative concentration. Nevertheless, the two types of measures share a trend: as in the case of relative concentration, we detect a falling tendency of the within-country share for a majority of sectors. The 1990s, however, are characterised by an apparent reversal in this tendency, that is by an increase in the within-country share of topographic concentration. That reversal is most manifestly evident for the manufacturing sector.

3.4.2 Factor decomposition

In Figure 7, we report proportional factor contributions (s_f) based on a decomposition of the topographic concentration of total employment using the GE(2) index (equation (8)) and the decomposition rule of equation (6). The manufacturing sector accounted for a continuously decreasing contribution to the topographic concentration of total employment. This result is consistent with the declining share of manufacturing in total employment (Figure 1) and its decreasing topographic concentration (Figure 5) - two factors which correspond to the second and third term respectively in the "natural" decomposition rule expressed by equation (6).

The factor-decomposition analysis also shows that non-market services on average account for the largest share of total topographic concentration. Hence, public-sector employment appears as the biggest contributor to the uneven geographical spread of economic activity.

4 Centre-periphery gradients and EU membership

The measures of geographic concentration used above possess the feature called "anonymity" in the income-distribution literature. Anonymity refers to the axiom that any permutation of basic units which changes only their ordering should not affect measured inequality. In other words, no attribute of a basic unit should matter except for its level of y. In the spatial context, this implies that no account is taken of the position of basic units (and observed units) relative to each other and relative to some fixed spatial reference point. In this section, we break the spatial anonymity inherent in the analysis of the previous section by identifying regions (i) according to their market potential, and (ii) by whether or not they belong to an EU member country. All results reported in this section are calculated from our data set 1.

4.1 The importance of being central

One of the principal insights of the "new economic geography" is that a location's market access can be a powerful attractor for increasing-returns activities.[21] The policy relevance of this issue is obvious.

4.1.1 The regression model

We use the peripherality index calculated by Copus (1999), which corresponds to the inverse of Harris's well-known market-potential measure:

$$P_r = \left(\sum_{s=1}^{R} \frac{G_s}{d_{rs}} \right)^{-1}, \tag{9}$$

where G_r denotes regional GDP in purchasing-power-parity terms, as computed by Eurostat; and d_{rs} stands for the distance in terms of road-freight travel time between regions r and s. Intra-regional distances d_{rr} are defined as one third of the longer axis of a rectangle bounding that region with north-south-east-west orientation.[22]

[21] In those models, the arrival of increasing-returns firms in a location is typically of sufficient magnitude that it increases the market potential of that location significantly and thereby triggers further arrivals of firms in a process of cumulative causation. Market access therefore becomes an endogenous variable. Our analysis abstracts from such processes by taking the market potential of regions as exogenous and time invariant. Our finding that the topographic concentration of overall employment has remained virtually unchanged over our sample period (see Figure 5) would seem to justify this restriction.

[22] Copus (1999) reports the peripherality measures in an interval ranging from 0, for the most central region (Paris), to 100, for the most peripheral (Northern Norway). Note that the appropriate measurement of distance, particularly at the intra-region level, continues to be the subject of lively debate (see, e.g., Head and Mayer, 2002). It would be interesting to test the sensitivity of our findings to alternative underlying distance measures, but such an exercise lies beyond the scope of this study.

Based on this measure, we compute centre-periphery gradients of our sample sectors by estimating the following simple specification separately for each sector and year:

$$\ln\left(\frac{\frac{y_{rf}}{\sum_f y_{rf}}}{\frac{\sum_r y_{rf}}{\sum_f \sum_r y_{rf}}}\right)_{rft} = \alpha_{ft} + \beta_{ft} P_r + \varepsilon_{rft}, \tag{10}$$

where, as before, y is employment and f stands for sectors. In addition, t denotes years, α and β are regression coefficients, and ε is a stochastic error. Our dependent variable is the log of what is commonly referred to as a Balassa index or location quotient (Overman *et al.*, 2001). This index scales sectoral employment by total employment, and hence it belongs to the class of *relative* concentration measures. We take logs in order to make the index symmetric.

Since there is evidence of between-country heteroskedasticity, we base our inference on White-adjusted t-statistics. To assess the statistical significance of changes in $\widehat{\beta}$ between sample years, we compute F tests on the hypothesis that $\widehat{\beta}_t - \widehat{\beta}_{t-x} = 0$, using seemingly unrelated regression estimates of the disturbance covariances in order to account for cross-equation error correlation (Greene, 2000: 620).

4.1.2 Results: centre-periphery gradients in Europe

Table 4 reports our results, based on sector-level regressions for 1975, 1987 and 2000. The results broadly conform with expectations based on casual observation. Agriculture is the only sector that exhibits a consistently positive and statistically significance locational bias towards peripheral regions. Conversely, three sectors are statistically significantly concentrated in central regions for all three sample years: manufacturing and energy, banking and insurance, and "other market services".

Looking at changes over time, we find that "other market services" is the only sector that exhibits a significant increase over the sample period in its tendency to concentrate at the centre. Conversely, three sectors have relocated significantly towards peripheral regions: manufacturing, construction and non-market services.

In the previous section we saw that the geographic distribution of manufacturing employment, conditioned on the distribution of total employment, has become tighter, while, conditioned on physical area, it has become more dispersed. Here we find that the centre-periphery dimension has lost some of its importance in shaping this distribution. We therefore conclude that manufacturing activity has been relocating away from high-density central regions.

4.2 The importance of being an EU member

One issue of particular interest from a policy perspective is the impact of EU integration on geographic concentration patterns. Exploiting the richness of our data set in terms of time coverage and intra-country information, we explore two questions: was accession to the EU associated with a change in the time profile of geographic concentration within countries? and: was accession to the EU associated with a change in the time profile of sectoral centre-periphery location trends?

4.2.1 The regression model

Neither theory nor prior empirical work give us much guidance on the specification of an econometric model that explains intertemporal changes in the geographic concentration patterns of individual sectors. Hence, we estimate a separate intercept and linear time trend for each country-sector over our full sample period, attributing to these intercepts and time trends all the forces that shape sectoral location patterns except for EU membership. Then, we estimate the deviation from this overall time trend of a time trend starting in the year of the relevant country's accession to the EU. Any deviation of the post-EU trend from the full-period trend is then interpreted as a membership effect. In order to obtain sufficient degrees of freedom for meaningful statistical analysis, and assuming that similar spatial forces were triggered when successive countries joined the EU, we force those deviation terms to be identical across countries and therefore estimate a unique membership effect per sector.

Specifically, we estimate the following regression model separately for each sector:

$$\mathbf{Z} = \mathbf{I}\boldsymbol{\alpha} + \mathbf{T}\boldsymbol{\beta} + \mathbf{E}\gamma + \boldsymbol{\epsilon}, \tag{11}$$

where

- K denotes the number of sample countries and T the number of sample years,
- $\mathbf{Z}$ is a $KT \times 1$ vector either of
  - within-country Theil indices of relative concentration, or of
  - estimated within-country centre-periphery gradients $\hat{\beta}$ from equation (10), regressed country-by-country;
- $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $K \times 1$ vectors of regression coefficients
- $\mathbf{I}$ is a $KT \times K$ matrix that consists of K diagonally stacked $T \times 1$ vectors of 1s, and zeros elsewhere;
- $\mathbf{T}$ is a $KT \times K$ vector consisting of K diagonally stacked $T \times 1$ vectors of sample years in ascending order ([1975, 1976,...,2000]) and zeros elsewhere;

- $\mathbf{E}$ is a $KT \times 1$ vector whose values are equal to the number of years either since the relevant country's accession to the EU or since 1975, whichever of the two is more recent, and zero for non-EU country-years;[23]

- γ is a regression coefficient (1×1); and

- $\boldsymbol{\epsilon}$ is a $KT \times 1$ vector of stochastic disturbances.

This is a piecewise linear spline function. The main object of our interest is the membership effect γ, a slope shifter contingent on accession to the EU.

In order to estimate equation (11), we need to take account of some dependencies in the data. Specifically, inspection of the data reveals significant intra-country autocorrelation and cross-country error correlation. Since the number of panels is relatively small ($K = 17$), we follow Beck and Katz (1995) and estimate the coefficients with feasible generalised least squares accounting for the intra-country autocorrelation (Prais-Winsten method) whilst taking account of the cross-country correlation and implied heteroskedasticity by basing inference on panel-corrected standard errors.

4.2.2 Accession to the EU and intra-country geographic concentration

The estimation results for the model with $\mathbf{Z}$ defined as within-country indices of relative concentration are reported in Table 5. For presentational reasons, we report only $\hat{\boldsymbol{\alpha}}$ and $\hat{\gamma}$.[24] Our model accounts for between 74% and 99% of the variance in the dependent variable. Accession to the EU has significantly affected within-country geographic concentration in three sectors: manufacturing, market services and non-market services. In all of these cases, EU accession has increased the slope of within-country concentration relative to time, hence, EU membership has been associated with increasing intra-country concentration of those three sectors. The within-country concentration of agriculture and construction, however, has not been affected by accession to the EU in a statistically significant way.

4.2.3 Accession to the EU and intra-country centre-periphery gradients

The results of the same exercise but with $\mathbf{Z}$ defined as estimated within-country centre-periphery gradients are reported in Table 6.[25] Again, our model accounts for most of the sample variance in the dependent variable, between 47% and 99%.

[23] We have experimented with alternative definitions of this varible, by starting the counter one or two years ahead of countries' accession dates, in order to take account of anticipatory relocation decisions. This made no qualitative difference to our results. The results are available upon request.

[24] In this section, we have amalgamated the market-services sectors into a single sector. Luxemburg had to be dropped from the data set, because for the intra-country concentration index to be computable, at least two regions are needed.

[25] Luxemburg and Ireland had to be dropped from the data set, because for the intra-country $\widehat{\beta}$ to be computable, at least three regions are needed.

The coefficient on the slope-shifting EU-accession variable is statistically significant in two sectors: manufacturing and market services. Accession to the EU is associated with an increasing tendency for manufacturing activity to locate in countries' peripheral regions. The opposite appears for market services, where EU accession is associated with an increasing tendency towards location in central regions.

5 Conclusions

We have provided an account of geographic concentration patterns in a broad range of sectors across Western European regions and countries from 1975 to 2000. Geographic concentration is quantified using entropy indices. These indices present two major advantages: they are decomposable, and they lend themselves to statistical inference through bootstrap tests. We distinguish between "relative" concentration, where location patterns are expressed relative to the spatial distribution of aggregate economic activity, and "topographic" concentration, where location patterns are expressed relative to physical space. In addition, we have estimated centre-periphery gradients in sectoral location patterns and assessed the impact of countries' accession to the European Union on changes in their internal economic geography.

Our study confirms the prevailing view of a European manufacturing sector that is slowly becoming more geographically concentrated, relative to the spatial spread of total employment (but not relative to physical space). We find that this process is statistically significant. Accession to the EU has strengthened concentration tendencies inside the new member countries. However, manufacturing concentration was not biased towards centrally located regions. The tendency of manufacturing activity to locate in economically central European regions has been significantly reduced over our sample period, and accession to the EU has strengthened those centrifugal location forces inside of the countries concerned. Finally, on manufacturing, we find a non-monotonic evolution of the within-country share in total geographic concentration, with a decrease in the 1970s and 1980s and an increase in the 1990s.

Service sectors are generally less geographically concentrated than manufacturing and agriculture. Market services have been re-locating towards economically central regions, and this process seems to have been reinforced by EU integration. Conversely, non-market services became increasingly located in peripheral regions, but this tendency did not appear to be influenced by EU integration.

The main aims of this paper were to propose more versatile measures for the description of geographic concentration patterns than those that have been used in most of this literature to date, and to provide a characterisation of locational trends in Western Europe. We believe that a rigorous and detailed description of changing concentration patterns is of interest in itself. Nonetheless, the value of such essentially descriptive results is of course enhanced if they can shed light on particular empirical, theoretical or policy-related questions.

When comparing our findings to those of related empirical studies, we essentially corroborate the prevailing view that sectoral location patterns in the EU are changing slowly at best, and that geographic concentration of manufacturing has been rising. Ours is the first study, to our knowledge, that underpins this result with statistical significance tests. Given that the geographic concentration of sectors and the sectoral specialisation of countries (or regions) are closely related concepts, we can also interpret our results as conforming with the remarkably robust empirical regularity found by Imbs and Wacziarg (2003), whereby countries' sectoral specialisation increases in income per capita once countries have crossed a certain GDP threshold. However, since we find that, in most sectors, observed concentration trends were increasingly driven by between-country changes (see Figure 3) our analysis casts doubt on whether their finding extends to the regional level.

A recently initiated and promising empirical approach is to estimate the relationship between geographic concentration and productivity. For Europe, Ciccone (2002) estimated the extent to which the topographic concentration of total employment increases regional labour productivity, which he called "agglomeration effects". Using regional cross-section data sets for the five largest West European countries, he found that productivity significantly rises in topographic concentration, with a remarkably robust elasticity of around 4.5 percent. Our analysis shows no significant change in the topographic concentration of total employment over time, but statistically significant changes in the topographic concentration of individual sectors (Table 3). The results of this paper therefore suggest that for an analysis of intertemporal agglomeration effects, sectorally disaggregated data would be more likely to yield significant results than the aggregate data used in Ciccone's cross-section estimations.

Aspects of our analysis can furthermore lend themselves to meaningful interpretation against trade and location theory. The decomposition of changes in geographic concentration into within- and between-country components holds particular promise in this respect. Venables (1999) has shown, based on a standard 2-country, 2-sector, 2-factor framework, that countries' sectoral specialisation will tend to increase as product-market integration proceeds, irrespective of whether the model is purely neoclassical (Heckscher-Ohlin) or whether it features "new economic geography"-type market-size effects. Yet, he also demonstrated that, in a neoclassical model, the specialisation effects of trade integration will be stronger if factor mobility is low, whereas in a "new economic geography" framework, specialisation effects are larger if factor mobility is high.[26] Given that in Western Europe factors continue to be more geographically mobile within countries than

[26] On the complementarity of factor mobility and geographic concentration forces (termed "agglomeration" in this context) in a new economic geography model featuring mobile firms as well as mobile labour, see Puga (1999).

between them, the decomposition of concentration trends might be interpreted as an informal test of competing theoretical paradigms. If increasing between-country concentration were indeed driven by neoclassical determinants, and an increase in the share of within-country concentration reflects market-size ("agglomeration") effects, then our findings would for example support the conjecture that, up to the 1980s, neoclassical factors have dominated the relocation of European manufacturing employment, whereas agglomeration forces have come to dominate since the 1990s. We believe that this link between theoretical predictions and the decompositions of entropy concentration measures would deserve further scrutiny.

With respect to the policy-related motivation for distinguishing within-country from between-country concentration, as discussed in the Introduction, we find that the average share of between-country concentration in total geographic concentration has been increasing. This suggests that shocks which originate in specific industries may be increasingly translating into country-specific shocks. On the question of the optimal jurisdictional level at which to conduct regional policy, our finding would favour delegation to supranational authorities.

We hope to have shown that, using appropriate quantitative techniques and sufficiently comprehensive data sets, descriptive empirics on economic geography can be a fruitful exercise. Yet, it is in the nature of such work that further improvements are not difficult to conceive. For example, the measure of centrality/peripherality could be made time-variant and sector-specific; and for even closer correspondence to location theory, one could separately estimate a region's access to input and output markets for each industry. It could also be interesting to describe evolutions of the full distribution of sectoral location patterns including transitions over time of region-sector observations inside those distributions, and to compute measures of spatial separation so as to assess the contiguity of sectoral clusters.

The biggest constraint on the quality of research on location patterns in Europe, however, is the quality of available sub-national data. Our analysis cannot entirely escape the spatial and sectoral aggregation biases inherent in conventional regional statistics, even though we do our best to minimise their distorting impact. If it were possible to merge plant-level micro-geographic data sets that have been collected in several European countries, ideally encompassing services as well as manufacturing establishments, the description of the European economic geography could take a quantum leap in terms of accuracy, detail and inference. In the meantime, we believe that our approach helps extract maximum information from the available statistical material.

References

[1] Aiginger, Karl and Steve W. Davies (2001) Industrial specialisation and geographic concentration: Two sides of the same coin? *Working Paper #23*, University of Linz.

[2] Aiginger, Karl and Wolfgang Leitner (2002) Regional concentration in the United States and Europe: who follows whom? *Weltwirtschaftliches Archiv*, 138(4): 652-679.

[3] Aiginger, Karl and Michael Pfaffermayr (2003) The Single Market and geographic concentration in Europe. *Review of International Economics*, forthcoming.

[4] Amiti, Mary (1999) Specialisation patterns in Europe. *Weltwirtschaftliches Archiv*, 135: 573-593.

[5] Arbia, Giuseppe (1989) *Spatial data configuration in statistical analysis of regional economic and related problems*. Kluwer.

[6] Beck, Nathaniel and Jonathan N. Katz (1995) What to do (and not to do) with time-series cross-section data. *American Political Science Review*, 89(3): 634-647.

[7] Biewen, Martin (2001) Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics*, 108: 317-342.

[8] Bourguignon, François (1979) Decomposable income inequality measures. *Econometrica*, 47: 901-920.

[9] Bourguignon, François and Christian Morrisson (2002) Inequality among world citizens: 1820-1992. *American Economic Review*, 92(4): 727-744.

[10] Brülhart, Marius (2001a) Evolving geographic concentration of European manufacturing industries. *Weltwirtschaftliches Archiv*, 137: 215-243.

[11] Brülhart, Marius (2001b) Growing alike or growing apart? Industrial specialisation of EU countries. In: Wyplosz, C. (ed.) *The Impact of EMU on Europe and the Developing Countries*, Oxford University Press.

[12] Ciccone, Antonio (2002) Agglomeration effects in Europe. *European Economic Review*, 46: 213-227.

[13] Clark, Todd and Eric van Wincoop (2001) Borders and business cycles. *Journal of International Economics*, 55: 59-85.

[14] Combes, Pierre-Philippe and Henry G. Overman (2003) The spatial distribution of economic activities in the European Union. In: Henderson V. and J.-F.

Thisse (eds.) *Handbook of Urban and Regional Economics*, Elsevier-North Holland, Amsterdam, forthcoming.

[15] Copus, Andrew K. (1999) A new peripherality index for the NUTS III regions of the European Union. Report for the Regional Policy Directorate of the European Commission.

[16] Cowell, Frank A. (1980) On the structure of additive inequality measures. *Review of Economic Studies*, 47: 521-531.

[17] Duranton, Gilles and Henry G. Overman (2002) Testing for localisation using micro-geographic data. *CEPR Discussion Paper* #3379.

[18] Duro, Juan A. and Joan Esteban (1998) Factor decomposition of cross-country income inequality, 1960-1990. *Economics Letters*, 60: 269-275.

[19] Fujita, Masahisa; Paul Krugman and Anthony J. Venables (1999) *The spatial economy: Cities, regions and international trade*, MIT Press.

[20] Fujita, Masahisa and Jacques-François Thisse (2002) *Economics of agglomeration*, Cambridge University Press.

[21] Giannetti, Mariassunta (2002) The effects of integration on regional disparities: Convergence, divergence or both? *European Economic Review*, 46: 539-567.

[22] Greene, William H. (2000) *Econometric analysis*, 4th ed., Prentice Hall.

[23] Haaland, Jan I.; Hans J. Kind; Karen H. Midelfart Knarvik; and Johan Torstensson (1999) What determines the economic geography of Europe?" *CEPR Discussion Paper* #2072.

[24] Haddad, Lawrence and Ravi Kanbur (1990) How serious is the neglect of intra-household inequality? *Economic Journal*, 100: 866-881.

[25] Hallet, Martin (2000) Regional specialisation and concentration in the EU. European Commission Economic Papers, No. 141.

[26] Head, Keith and Thierry Mayer (2002) Illusory border effect: distance mismeasurement inflates estimates of home bias. *CEPII Working Paper* #2002-1, Paris.

[27] Helg, Rodolfo, Paolo Manasse, Tommaso Monocelli and Riccardo Rovelli (1995) How much (a)symmetry in Europe? Evidence from industrial sectors. *European Economic Review*, 39: 1017-1041.

[28] Imbs, Jean and Romain Wacziarg (2003) Stages of diversification. *American Economic Review*, 93: 63-86.

[29] Kalemli-Ozcan, Sebnem; Bent E. Sorensen and Oved Yosha (2003) Risk sharing and industrial specialization: regional and international evidence. *American Economic Review*, forthcoming.

[30] Krugman, Paul (1991) *Geography and Trade.* MIT Press.

[31] Midelfart Knarvik, Karen H.; Henry G. Overman; Stephen J. Redding and Anthony J. Venables (2000) The location of European industry. European Commission Economic Papers, No. 142.

[32] Mills, Jeffrey A. and Sorushe Zandvakili (1997) Statistical inference via bootstrapping for measures of inequality. *Journal of Applied Econometrics*, 12: 133-150.

[33] OECD (2001) *OECD territorial outlook.* Paris.

[34] Overman, Henry G.; Stephen Redding and Anthony J. Venables (2001) The economic geography of trade, production and income: a survey of empirics. *CEPR Discussion Paper* #2978.

[35] Peri, Giovanni (1998) Technological growth and economic geography. *IGIER Working Paper* #139, Bocconi University.

[36] Quah, Danny (1996) Regional convergence clusters across Europe. *European Economic Review*, 40: 951-958.

[37] Puga, Diego (1999) The rise and fall of regional inequalities. *European Economic Review*, 43: 303-334.

[38] Shorrocks, Anthony F. (1980) The class of additively decomposable inequality measures. *Econometrica*, 48: 613-625.

[39] Shorrocks, Anthony F. (1982) Inequality decomposition by factor components. *Econometrica*, 50: 193-211.

[40] Shorrocks, Anthony F. (1984) Inequality decomposition by population subgroups. *Econometrica*, 52: 1369-1385.

[41] Slottje, Daniel J. (1990) Using grouped data for constructing inequality indices: Parametric vs. nonparametric methods. *Economics Letters*, 32: 193-197.

[42] Storper, Michael; Yun-chung Chen and Fernando De Paolis (2002) Trade and the location of industries in the OECD and European Union. *Journal of Economic Geography*, 2: 73-107.

[43] Venables, Anthony J. (1999) Trade liberalisation and factor mobility: An overview. In: Riccardo Faini et al. (eds.) *Migration: The controversies and the evidence*, Cambridge University Press.

Table 1: Relative concentration of sectors, 1975-2000 [1]

(employment, 236 regions)

Sector	**Avg** $\boldsymbol{GE(1)}$[2]	$\boldsymbol{\Delta GE(1)_{75-00}}$	$\boldsymbol{\Delta GE(1)_{75-87}}$	$\boldsymbol{\Delta GE(1)_{87-00}}$	**Share**[3]
Agriculture	0.474	0.029	0.008	0.021	0.07
Manufact., energy	0.055	0.020**	0.004	0.016**	0.24
Banking, insurance	0.053	0.004	-0.012	0.016	0.04
Non-mkt services	0.041	-0.023	-0.022	-0.001	0.22
Transport, communic.	0.036	-0.043**	-0.036**	-0.007*	0.05
Distributn	0.031	0.007	0.002	0.004	0.13
Other mkt services	0.030	-0.005	-0.008	0.003	0.16
Constructn	0.025	0.019	-0.012*	0.031**	0.07

[1] **/* denotes rejection of H0 that $\Delta GE(1) = 0$, based on bootstrap 95%/90% confidence intervals (10,000 replications)

[2] Average annual $GE(1)$ index (employment weighted) over 1975-2000 period

[3] Sector share in total employment over the full sample period

Table 2: Relative concentration of manufacturing sectors, 1980-1995 [1]
(gross value added, 116 regions)

Sector	**Avg** $GE(\mathbf{1})^2$	$\mathbf{\Delta}GE(\mathbf{1})_{\mathbf{80-95}}$	$\mathbf{\Delta}GE(\mathbf{1})_{\mathbf{80-87}}$	$\mathbf{\Delta}GE(\mathbf{1})_{\mathbf{87-95}}$	**Share**[3]
Ores, metals	0.389	-0.0555	-0.0551*	-0.0004	0.04
Textiles, cloth., footw.	0.379	0.1649**	0.0534**	0.1115**	0.08
Transport eq.	0.163	0.0196	0.0216	-0.0020	0.10
Chemicals	0.152	0.0003	0.0085	-0.0082	0.10
Non-metallic minerals	0.142	0.0171	0.0016	0.0156	0.06
Misc. manuf.	0.111	-0.0044	-0.0064	0.0020	0.09
Machinery, electronics	0.109	-0.0057	0.0180**	-0.0238**	0.31
Paper prod.	0.104	0.0098	-0.0022	0.0120	0.08
Food, drink, tobacco	0.082	0.0114	0.0026	0.0088	0.13
Tot. manuf.	*0.043*	0.0023	*0.0071***	*-0.0048*	*1.00*

[1] **/* denotes rejection of H0 that $\Delta GE(1) = 0$, based on bootstrap 95%/ 90% confidence intervals (10,000 replications)
[2] Average annual $GE(1)$ index (GVA weighted) over 1980-1995 period
[3] Sector share in total employment over the full sample period

Table 3: Topographic concentration of sectors, 1975-2000
(employment, 236 regions)

Sector	**Avg** $GE(\mathbf{1})^1$	$\mathbf{\Delta}GE(\mathbf{1})_{\mathbf{75-00}}$ [2]
Other market services	1.039	-0.016
Transport, communication	1.028	-0.148**
Banking, insurance	1.008	-0.024
Distribution	0.938	-0.052
Non-market services	0.890	-0.140*
Manufacturing, energy	0.868	-0.161**
Construction	0.738	0.008
Agriculture	0.490	0.104**
Total employment	*0.810*	*-0.002*

[1] Average annual $GE(1)$ index (area weighted), 1975-2000
[2] **/* denotes rejection of H0 that $\Delta GE(1) = 0$, based on bootstrap 95%/90% confidence intervals (10,000 replications)

Table 4: Centre-periphery gradients, 1975-2000[1] (236 regions)

Sector	**Year**	$\hat{\beta} * 100$	R-sq	$(F \mid H0\text{: } \hat{\beta}_t\text{-}\hat{\beta}_{t-x}\text{=}0)^2$	
				$x \in \{12, 13\}$	$x = 25$
Agriculture	*1975*	3.46**	0.29		
	1987	3.30**	0.29	3.2	
	2000	3.36**	0.27	0.1	0.1
Manufacturing, energy	*1975*	-1.05**	0.23		
	1987	-0.87**	0.17	6.5*	
	2000	-0.47**	0.05	55.7**	32.6**
Construction	*1975*	-0.19	0.01		
	1987	0.15*	0.02	12.5**	
	2000	0.36*	0.04	4.6*	18.0**
Distribution	*1975*	-0.28**	0.04		
	1987	-0.07	0.00	10.5**	
	2000	-0.16	0.01	3.8	1.7
Transport, communications	*1975*	-0.01	0.00		
	1987	-0.10	0.00	1.1	
	2000	-0.21*	0.02	5.6*	3.3
Banking, insurance	*1975*	-1.13**	0.19		
	1987	-1.09**	0.24	0.2	
	2000	-1.15**	0.20	0.5	0.1
Other market services	*1975*	-0.55**	0.10		
	1987	-0.64**	0.18	1.8	
	2000	-0.84**	0.30	13.1**	9.1**
Non-market services	*1975*	-0.27	0.01		
	1987	-0.11	0.00	7.2**	
	2000	0.30*	0.03	96.9**	50.5**

[1] see eq. (10); **/* denote stat. significance at 99%/95%, White-corrected

[2] F-statistic on Wald test of equality of $\hat{\beta}$ across years, taking account of cross-equation error covariance

Table 5: EU membership and intra-country relative concentration					
	Dependent variable = intra-country Theil index (employment, 16 countries) (reported coeff. = estimated coeff. * 100):				
Indep. vars:	**Agric.**	**Manuf.**	**Constr.**	**Mkt serv.**	**Non-mkt s.**
Fixed effect:					
Belgium	45.3*	9.5*	5.4	24.4*	11.3*
Denmark	39.3*	3.6	18.3*	21.5*	21.8*
Germany	16.2	7.5*	7.7	22.6*	24.4*
Greece	-24.3	47.5*	35.1*	35.4*	55.3*
Spain	3.7	20.0*	17.6*	18.0*	27.7*
France	24.2	12.2*	19.9*	24.2*	20.8*
Ireland	-1.0	14.0*	17.4*	20.9*	21.0*
Italy	-30.4	21.5*	23.1*	22.9*	25.9*
Netherlands	6.2	6.7	14.4*	26.5*	20.8*
Austria	31.9*	-2.3*	7.6*	10.5*	9.3*
Portugal	19.1	-4.0	14.9*	57.1*	30.8*
Finland	-0.7	6.5*	4.9*	9.1*	7.6*
Sweden	21.5*	2.6	5.8*	5.4	7.6*
UK	91.5*	9.1*	16.5*	21.9*	21.6*
Norway	9.4	0.7	-4.1	2.9*	2.7*
Switzerland	9.1*	-2.9*	0.6*	1.3*	0.4*
EU effect	-0.09	0.17*	0.22	0.27*	0.27*
R^2	0.99	0.99	0.74	0.86	0.98
Notes: Prais-Winsten GLS regressions with panel-corrected standard errors, assuming contemporaneous cross-panel correlation; interactions of fixed effects with year variable included but not reported; 416 obs.; * denotes 99% statistical significance.					

Table 6: EU membership and intra-country C-P gradients					
	Dependent variable = $\widehat{\beta}_{ft}$ (employment, 16 countries) (reported coefficients = estimated coefficients * 100)				
Indep. vars:	**Agric.**	**Manuf.**	**Constr.**	**Mkt serv.**	**Non-mkt s.**
Fixed eff.:					
Belgium	32.3*	4.7	13.3	-32.7*	-14.6*
Denmark	23.9*	2.6	14.9*	-31.6*	-13.3
Germany	17.8*	2.5	5.0	-28.1*	-5.6
Greece	10.6	1.9	6.8	-17.8*	-3.6
Spain	8.9	4.9	16.0*	-17.1*	-8.1
France	8.2	4.9	16.6*	-30.4*	-8.7
Italy	-0.1	5.7	15.2*	-31.1*	-9.3
Netherlands	14.9*	8.5	8.5	-32.3*	-11.1
Austria	14.1*	1.4	5.5*	-25.5*	-2.6
Portugal	25.5*	13.2*	32.4*	-75.2*	-17.7*
Finland	19.1*	-16.6*	11.8	-12.0*	-16.2*
Sweden	4.0	-3.9*	7.5*	-4.2	-1.3
UK	13.0	8.3	11.9	-28.4*	-11.9
Norway	7.5*	-8.4*	12.1*	-2.4	-0.4
Switzerland	11.8*	-10.3*	1.0	0.6	-8.8*
EU effect	0.02	1.3*	0.2	-0.4*	-0.1
R^2	0.99	0.98	0.72	0.83	0.47

Notes: Prais-Winsten GLS regressions with panel-corrected standard errors (see Beck and Katz, 1995); interactions of fixed effects with year variable included but not reported; 390 observ.
* denotes 99% statistical significance.

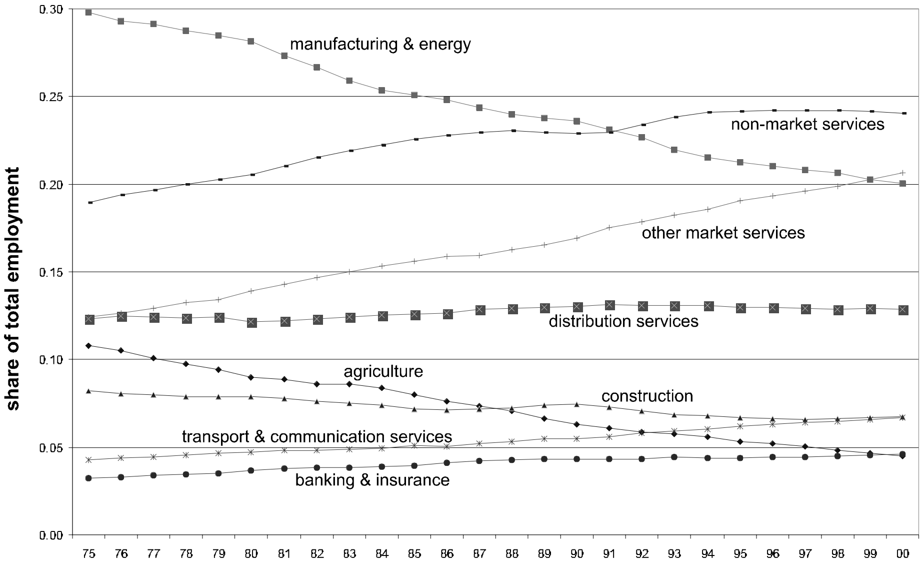

Figure 1: Sector shares in total employment, 1975-2000

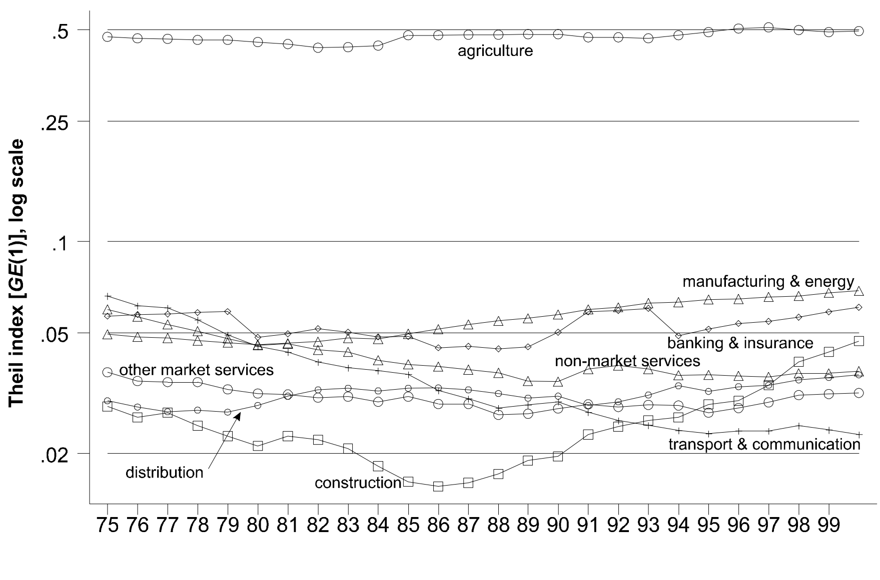

Figure 2: Relative concentration of sectors (Theil index, employment), 1975-2000

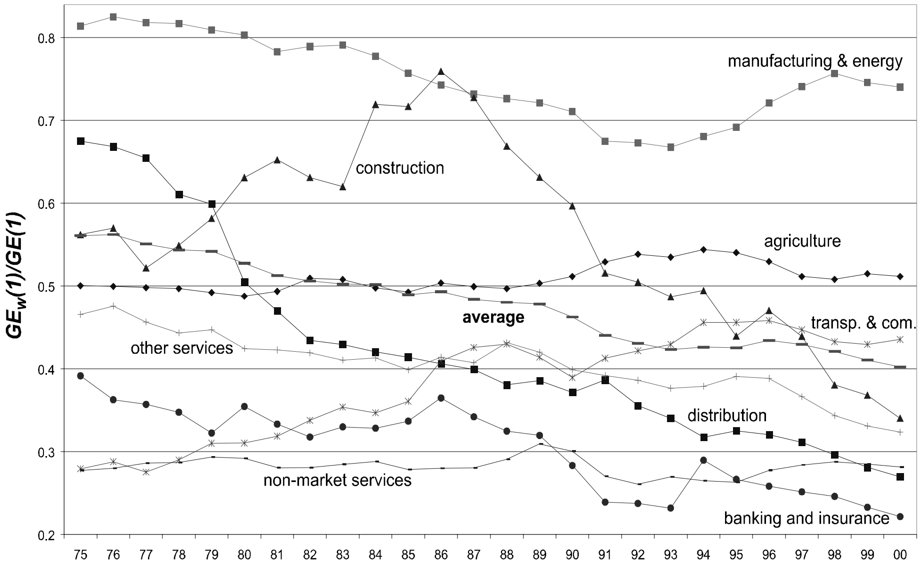

Figure 3: Within-country share in overall relative concentration (employment), 1975-2000

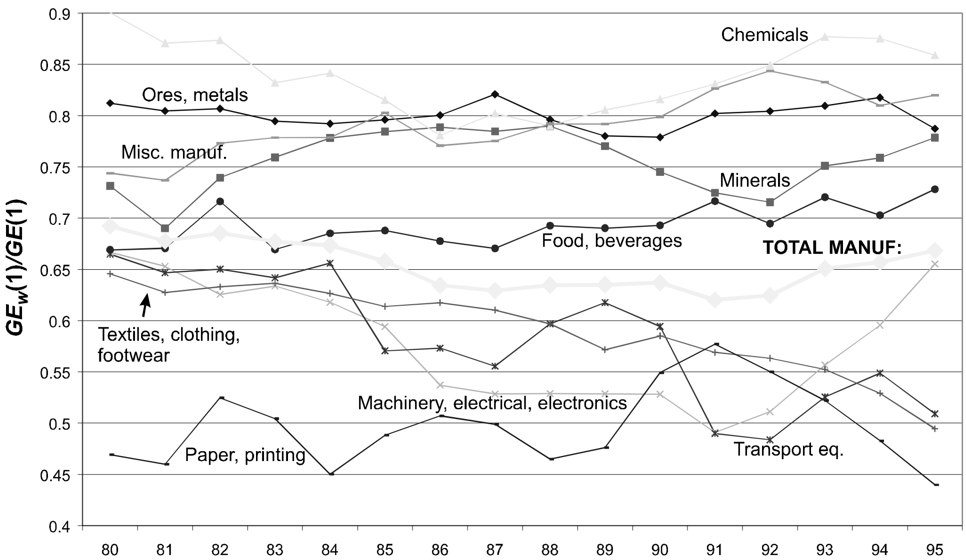

Figure 4: Within-country share in overall relative concentration of manufacturing sectors (GVA), 1980-1995

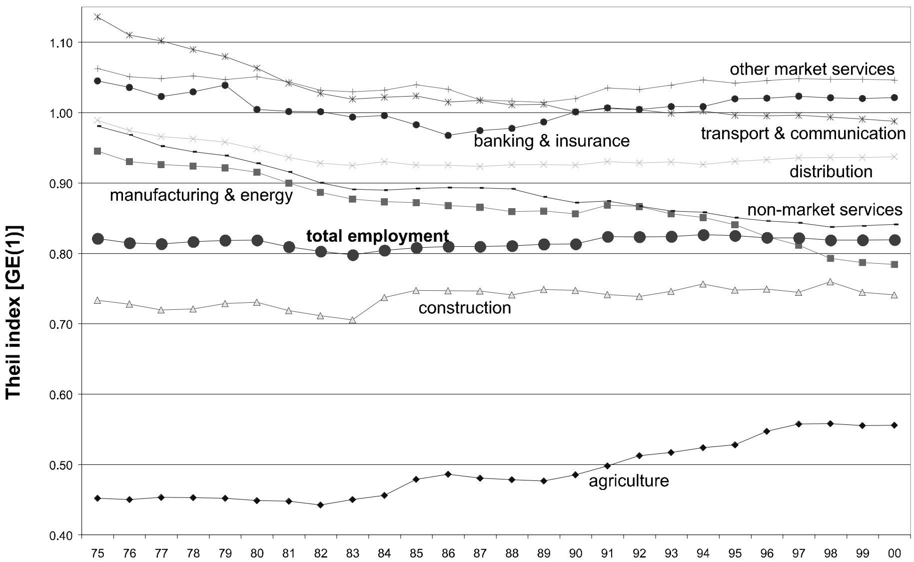

Figure 5: Topographic concentration of sectors (Theil index, employment), 1975-2000

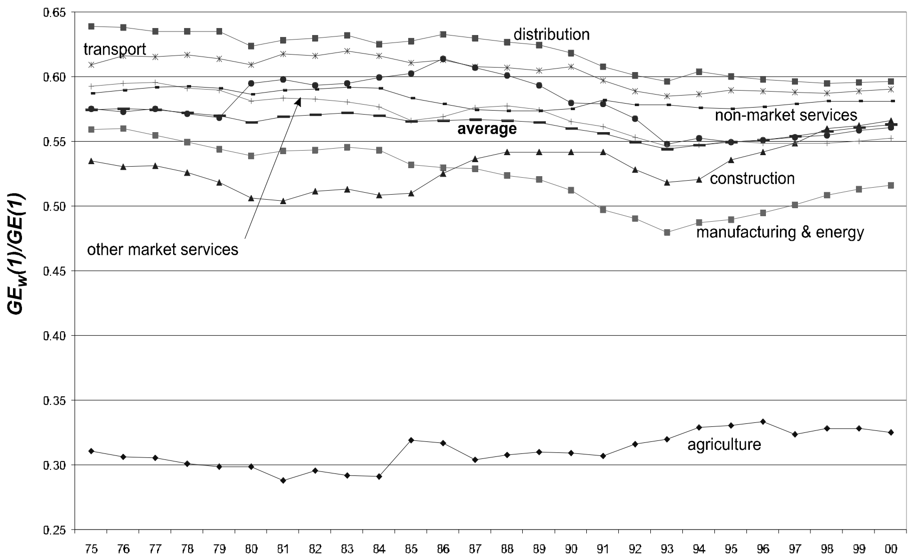

Figure 6: Within-country share in overall topographic concentration (employment), 1975-2000

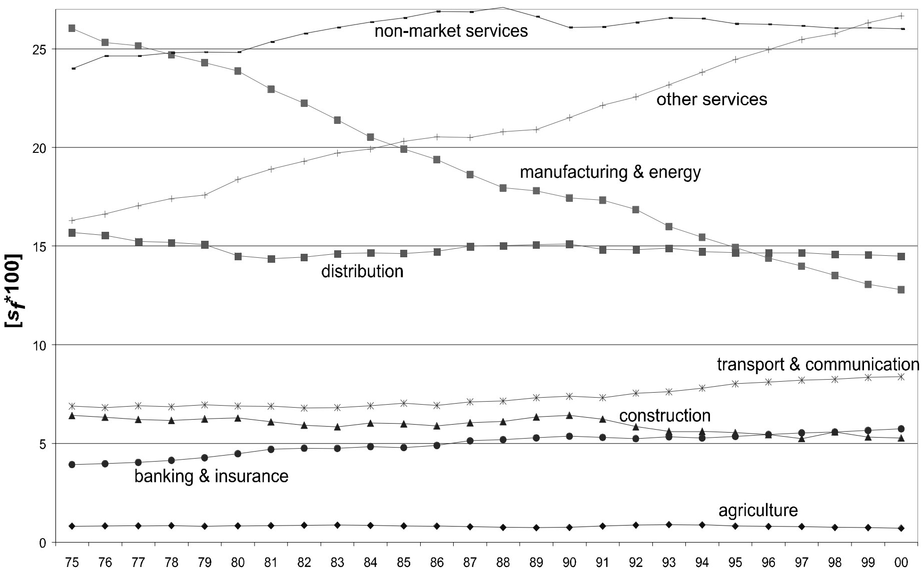

Figure 7: Sectoral "factor contributions" to topographic concentration (employment, GE(2) index), 1975-2000

A Appendix 1: Illustrations of geographic concentration indices

We provide here some examples of the changes in our indices for two simple scenarios of changing geographic concentration patterns. In both scenarios, we assume a universe of two observed units (i.e. regions), and we do not consider concentration patterns inside of those observed units. In scenario I, we assume that the two regions are identical in every respect bar their shares of Y (i.e. activity in the sector of interest). One can therefore abstract in this example from weighting issues, and treat the observed units as if they were basic units. We track the values of our measures as activity in the sector of interest changes from being fully concentrated in one region to being perfectly dispersed across the two regions. In scenario II, we assume that activity in the sector of interest remains equally split between the two regions, and we vary the underlying sizes of those regions instead. We can thus no longer treat regions, the observed units, as if they were the basic units. We track the values of our measures as the sizes of two regions move from being very unequal to being perfectly identical.

The two scenarios thus illustrate the two possibilities of changing geographic concentration of an individual sector: relocation of the sector of interest, or changes in the region sizes with unchanged location of the particular sector. Note that the two types of changes are not necessarily independent. If we compute measures of relative concentration of a large sector, then the geographic concentration of that sector will affect relative region sizes. We abstract from this issue here and assume the two components to be independent (which strictly applies to the case of topographic concentration and of relative concentration of infinitesimally small sectors).

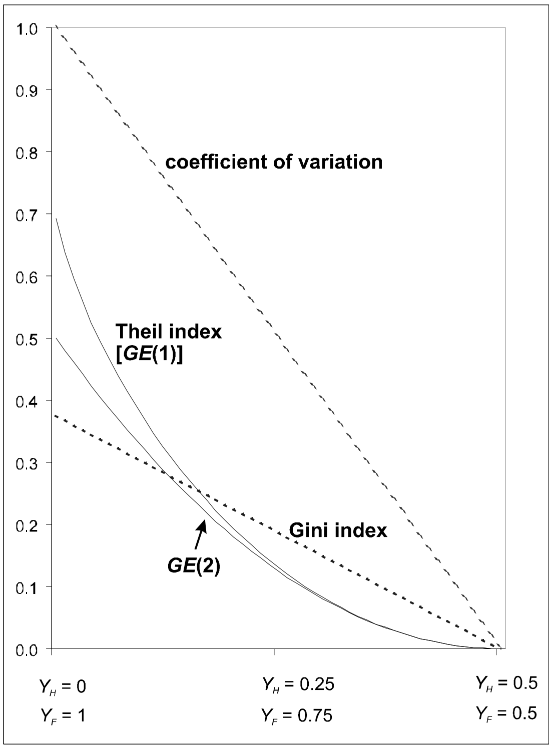

Figure A1: Sectoral relocation between two regions (Scenario 1)

Scenario 1: Suppose two identical regions, H and F. The world size of the sector, Y, is assumed constant and equal to 1, but its distribution across H and F is allowed to change. Moving from left to right in Figure A1, we start from a situation where all of that sector's activity is concentrated in region F, so that $Y_H = 0$, and then gradually move activity out of region F and into region H, until $Y_H = Y_F = 0.5$.

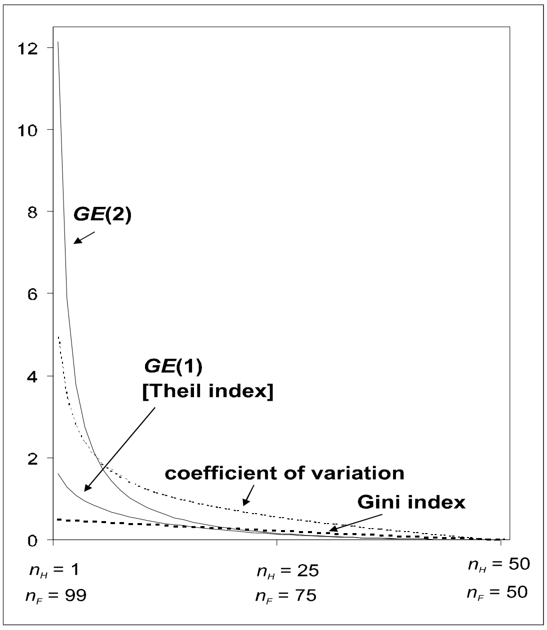

Figure A2: Unchanged sectoral location in two changing regions (Scenario 2)

Scenario 2: Suppose the two regions can have different sizes, n_H and n_F, but that $Y_H = Y_F = 0.5$ throughout. The size of the world is set to 100 ($N = n_H + n_F = 100$). Moving from left to right in Figure A2, we start from a situation with very unequally sized regions, where $n_H = 1$ and $n_F = 99$, and then gradually equalise region sizes, until $n_H = n_F = 50$.

Both our scenarios simulate a reduction in geographic concentration. The graphs show that our indices always fall in geographic concentration, and that the entropy indices are monotonic transformations of each other but neither of the Gini index nor of the coefficient of variation.

B Appendix 2: Data

B.1 Data set 1

- Source: Cambridge Econometrics Regional Database (based on Eurostat's REGIO and national sources)
- Variable: employment
- Time dimension: annual averages, 1975-2000
- Sectors: agriculture; manufacturing and energy; construction; distribution; transport and communications; banking and insurance; other market services; non-market services (8 sectors, based on NACE-CLIO classification)
- Regional breakdown: 236 regions, see Table A1
- Number of observations: 49,088

B.2 Data set 2

- Source: Hallet (2000) (based on Eurostat's REGIO and national sources)
- Variable: gross value added
- Time dimension: annual averages, 1980-1995
- Sectors retained: ores and metals; non-metallic minerals; chemicals; metal goods, machinery and electrical goods; transport equipment; food products; textiles, clothing and footwear; paper and printing products; misc. manufactured goods (9 industrial sectors, based on NACE-CLIO classification)
- Regional breakdown: 116 regions, see Table A1 (French "Départements d'outremer" as well as Madeira and Açores were dropped from Hallet's original data set, in order to enhance comparability with data set 1).
- Number of observations in full data set: 32,368

Data set 1					Data set 2				
Country	Number of regions for which data are available[1]	Administrative units	Classification level[2]	Observations	Country	Number of regions for which data are available[1]	Administrative units	Classification level[2]	Observations
Belgium	10	Provinces	NUTS 2	Vlaams Brabant and Brabant Wallon clustered as one region	Belgium	11	Provinces	NUTS 2	
Denmark	3	Regions	TL 2		Denmark	1			
Germany	31	Regierungsbezirke	NUTS 2	Neue Länder excluded	Germany	10	Länder	NUTS 1	Berlin and neue Länder excluded
Greece	13	Development regions	NUTS 2		Greece	1			
Spain	18	Comunidades autónomas + Ceuta y Melilla	NUTS 2		Spain	18	Comunidades autónomas + Ceuta y Melilla	NUTS 2	
France	22	Régions	NUTS 2	DOMs excluded	France	22	Régions	NUTS 2	DOMs excluded
Ireland	2	Regions	NUTS 2		Ireland	1			
Italy	20	Regioni	NUTS 2		Italy	20	Regioni	NUTS 2	
Luxembourg	1				Luxembourg	1			
Netherlands	12	Provincies	NUTS 2		Netherlands	12	Provincies	NUTS 2	
Austria	9	Bundesländer	NUTS 2		Austria	1			
Portugal	5	Comissões de coordenação regional	NUTS 2	Regiões autónomas excluded	Portugal	5	Comissões de coordenação regional	NUTS 2	Regiões autónomas excluded
Finland	6	Suuralueet	NUTS 2		Finland	1			
Sweden	21	Län	NUTS 3		Sweden	1			
United Kingdom	37	Counties or groups of unitary authorities	NUTS 2		United Kingdom	11	Government office regions	NUTS 1	According to NUTS 95 classification
Norway	19	Fylker	TL 3		Norway				
Switzerland	7	Grandes régions	TL 2		Switzerland				
TOTAL EU15	210				TOTAL EU15	116			
TOTAL WE17	236								

Table A1: Regional breakdown of data sets 1 and 2