Testing for localization with relative entropy measures

Roy Cerqueti
University of Macerata, Department of Economics and Law
Via Crescimbeni, 20. I-62100, Macerata, Italy
Email: roy.cerqueti@unimc.it

Eleonora Cutrini
University of Macerata, Department of Law
Piaggia dell'Università, 2. I-62100, Macerata, Italy
Email: eleonora.cutrini@unimc.it

Extended abstract

It is widely recognized that industrial clustering is a key stylized fact of economic geography. Firms tend to collocate because of exogenous natural advantages, knowledge spillovers, input-output linkages and/or labor market pooling.

During the last decade, the development of New Economic Geography models was inevitably accompanied by a rising interest in measuring localization of economic activities. Ideally, geographical concentration should be assessed in continuous space to avoid the checkerboard problem related to arbitrary spatial units (See Arbia, 2001 for details on aggregation and scale problems as two different manifestations of the arbitrariness of geographical boundaries). Despite these prominent advantages, distance-based measures may be not the "first-best" metrics, not only because they require micro-geographical data (which are still not always available), but also because, when geographical boundaries are expressions of meaningful and active economic institutions instead of artificial border bias, it may be important to know the spatial scale at which

clustering occurs to inform policy makers. This can be certainly appropriate for the European area where national sovereignty has not been completely discarded throughout the process of economic integration. Some authors propose the use of relative entropy indices (Mori *et al*., 2005; Brülhart and Traeger, 2005; Bickenbach and Bode, 2008, Cutrini, 2009, 2010), which have distinct advantages over the standard concentration measures constructed over the discrete space. The most relevant one is their decomposability. This feature allows authors to decompose the inequality analysis across different spatial and sectoral scale in order to identify the contributions of individual regions (sectors) to the overall localization of economic activity.

Moreover, relative entropy measures allows for an integrated analysis of localization where the complexity of the real location patterns are reduced to two characteristics dimensions (concentration/specialization), so that the spread of economic activities across space is mirrored by the structural differences between geographical units (equivalence between regional specialization and industrial concentration) (Cutrini, 2009).

Relative measures should be preferred over absolute measures to gauge and test for the spatial structure of economic activities (see also Lang *et al.*, 2014 on this view) as long as the aim of the analysis is to assess the strength of industry-specific localization economies. In fact, the two types of measures (absolute and relative) differ fundamentally from each other in terms of what is considered to represent no localization, or noconcentration. While for absolute measures, the reference is the uniform distribution, in the case of relative measure the regional distributions of an industry is compared to the regional distribution of total economic activity,. This is the logic underlying the Ellison and Glaeser index (Ellison and Glaeser, 1997) and the entropy measures of relative concentration and localization. Even in the context of distance-based methods, the main contributions have originally focused on absolute versions (Marcon and Puech, 2003; Duranton and Overman, 2005), but they have recently been refined to control for the distribution of the economic activity as a whole (Lang *et al.*, 2014).

Relative entropy measures are based on the most commonly used indicator to detect region-industry localization economies, the Location Quotients (LQs). Hence, it is also worth mentioning some recent attempts to link LQs to existing theoretical models of location. Guimarães *et al.* (2009) interpreted the LQ as derived from the stochastic model of firm location put forward by Ellison and Glaeser (1997). In this way, they show how to interprete the value of the statistics in relation to the strenghts of externalities or natural advantages and provides a theoretical foundation on which to build statistical tests of significance using established statistical theory.

Billings and Johnson (2012a) developed a test for detecting specialization, by constructing a bivariate kernel density estimator of establishment concentration within a given industry. They suggest that the global estimator proposed is unbiased for small samples in the creation of the counterfactual distribution. Similarly, the same authors, in a companion paper, while developing testing procedures based on this measure, suggest that the LQ is typically unbiased, but exhibits finite sample bias when assuming a Poisson distribution (Billings and Johnson, 2012b).

Before entering the details of our contributions, some further premises are needed.

This paper aims to add a statistical testing procedure to entropy-based measure of relative concentration. In this respect, it is in line with research on the simulation of confidence intervals for measures constructed over the discrete space such as the Ellison and Glaeser index to test the null hypothesis of no concentration (Cassey and Smith, 2014).

Our approach is also related to recent advancements in the context of distance-based methods for absolute indices (Duranton and Overman, 2005; Marcon and Puech, 2003, 2010) and relative indices (Lang *et al.*, 2014) in that they also seek to detect significant localization patterns over the continuous space.

In this paper, the testing procedure developed to evaluate whether observed spatial distribution of industries is significantly different from regularity by controlling for industrial structure.

Tests are specific for industrial sectors (concentration), and consider the situations of overall localization, spatial concentration within countries and between countries. The proposed approach is based on the presence of economic factors –specific for each industry.- leading to a not completely uniform distribution of manufacturing employment across regions. More precisely, we rely on the Ellison and Glaeser's approach (EG henceforth) to derive a meaningful null hypothesis of random choice. Then, in the same vein as in the EG approach, we take into account that, even in the case of random firms' location choice, the observed employment distribution will not be perfectly regular because of the lumpiness of establishments. In fact, it may well be that even when firms choose independently from each other their production sites, the spatial distribution of employment is not stochastic because of the industrial structure. Hence, a certain degree of "dissimilarity" of employment distribution of a manufacturing sector relative to the reference (e.g total manufacturing employment) will be perfectly consistent with the independent location model of firms.

The null hypotheses of the tests are defined by introducing critical thresholds, whose values are endogeneized in the application on the basis of empirical data and identified through a Montecarlo procedure. Specifically, the construction of counterfactual distributions for each industry is performed on a sample of NUTS2 regions for 16 countries and for the reference year 2007 (see section 4.1 for the details on the data). For each industrial sector, 10,000 resample of the available data are drawn on the basis of their empirical distribution. Then, the relative entropies of the resampled series are computed, and the 5% and 95% percentiles of the entropies distributions are assessed. This gives the critical values of the tests at the related confidence levels.

To the best of our knowledge, this is the first contribution dealing with the development of statistical tests based on relative entropies. The decomposition properties typical of entropy measures allows distinguishing the relevance and significance of agglomeration and dispersion at the different spatial scales (global vs local).

The paper is organized as follows: Section 1 reviews the related literature. Section 2 introduces the framework clarifying some relevant issues related to the selection of the appropriate variable of interest, its economic meaning, the construction of the reference distribution for relative measures, and the associated underlying random location model. Section 3 presents the testing procedure through Montecarlo simulations. Section 4 includes an application based on regional data. Section 5 offers some concluding remarks.