MEASURING AGGLOMERATION

SCOTT DUKE KOMINERS

1. INTRODUCTION

Since Marshall (1920), economists have recognized the propensity for industries to agglomerate across space.[1] This effect is not an accident—spatial clustering results in increased returns and growth, as a consequence of localized economies of scale.

The reduced transport costs within an agglomeration lead to "physical spillovers," as discussed by Krugman (1991); these spillover effects were found by Ciccone and Hall (1996) to be sufficient to offset congestion effects. Furthermore, as documented by Glaeser *et al.* (1992), close geographic proximity leads to "intellectual spillovers," localized information flows which increase firm productivity.

Despite the importance of firms' location decisions to eventual efficiency, the study of economic geography has been fairly neglected until recently. This omission of geographic considerations owed, primarily, to the difficulties inherent in constructing models which effectively interpret spatial dimensions.[2]

Within the past decade, however, the "New Economic Geography" literature begun by Krugman (1991) has indicated that spatial agglomeration is quite common. Indeed, it now seems apparent that industries are more likely to be agglomerated than they are to be dispersed.[3]

The importance of agglomeration varies across industries, whence different industries have been found to exhibit surprisingly different levels of agglomeration (see Ellison and Glaeser (1997)). A complete understanding of the pathways by which agglomeration increases firms' returns requires

---

This paper was originally composed for the Harvard Urban and Social Economics seminar (Economics 2800b). The author thanks Zachary Abel, Natalie Bau, Kjell Carlsson, Daniel Li Chen, Buck Farmer, Jocelyn Finlay, Mariel Finucane, Edward Glaeser, Elizabeth Hadaway, Brett Harrison, John Hawksley, Marie Herring, Gregory Ihrie, Paul Kominers, Menyoung Lee, Daniel Litt, Andrei Shleifer, Brett Simchowitz, Robert W. Sinnott, and Crystal Yang for helpful discussions and comments on this paper.

[1]For a comprehensive discussion of agglomeration effects, see Fujita and Thisse (2002).

[2]Indeed, much of the spatial equilibrium and economic geography literature based on Krugman (1991) relies on the somewhat aspatial assumption that firms locate in discrete "states." This is clearly not realistic, and some of the most recent literature which we will discuss has attempted to work in a more continuous framework. The generalization to continuous space is valid, in general, if difficult to study. In particular, Quah (2002) gives a highly technical theory justification for existence and dynamics of spatial agglomeration across 3-dimensional space.

[3]All of the agglomeration indices we will discuss allow us to draw this conclusion, which first appeared in Krugman (1991). There is some dispute about the overall ratio, however. For example, Ellison and Glaeser (1997) find that 97% of United States industries are concentrated, whereas Duranton and Overman (2002) argue that only 43% of industries are actually localized. This will be discussed in more detail in Section 2.2, below.

an effective *agglomeration index*, a measurement which identifies the concentration of industry and explains the relationship between firm location choice and industry concentration.

It has proven quite difficult, however, to find such a mesurement. The problem of measuring spatial concentration, alone, is difficult: Duranton and Overman (2002) argued that a satisfactory spatial agglomeration measurement should

(1) be comparable across industries,
(2) control for overall agglomeration trends across industries,
(3) separate spatial concentration from industrial concentration,
(4) be unbiased with respect to the degree of spatial aggregation, and
(5) admit a clear statistical significance test.

Further complicating the problem, effective agglomeration indices must, for practical reasons, be

(6) computable in closed-form from accessible data.

In addition, an index is almost meaningless if it is not

(7) justified by a suitable model,

as an index without a supporting model does not give rise to a realistic interpretation.

An index satisfying all seven of these conditions would allow economists to easily and effectively characterize the determinants of agglomeration across industries, in turn giving insight into the effects by which aggmomeration drives firm productivity. However, the difficulties inherent in the problem have, at present, proven prohibitive—no agglomeration index in the literature satisfies all of the conditions (1)-(7). The most axiomatically complete indices are those with the most complicated data requirements and weakest supporting interpretations. Thus, economists have not agreed upon which indices to use. Since agglomeration studies which use inconsistent indices are, as we shall see, rarely comparable, this inconsistency presents a serious problem to those hoping study the determinants and dynamics of agglomeration.

In this paper, we will discuss the current body of agglomeration measurement literature. In Section 2, we focus our attention on the variety of agglomeration indices which have been developed in response to Krugman (1991). We explain and critique both the discrete and continuous agglomeration measurement literature. We then, in Section 3, examine some of the studies which have applied and compared the different agglomeration indices. We also discuss the studies on general agglomeration dynamics, in Section 3.2. Section 4 concludes.

2. AGGLOMERATION INDICES

As the economic geography literature has evolved, there has been a divergence between model-based and axiomatic approaches to the agglomeration measurement question.

The model-based agglomeration indices have all relied on the assumption of discrete spatial units, that is, they assume that firms choose to locate in discrete "states" which are all equidistant from one another. Many of the agglomeration measurements which refine the model-based approaches also rely on this assumption, even if they do not provide an economic choice model which requires this approach.

A newer generation of indices, meanwhile, has done away with the assumption of discreteness. This literature has, instead, based its measurement of firm clustering on distance density measurements, which are independent of political boundaries.[4] This class of measurements has its own problems, however, as such measurements are hard to interpret and very difficult to calculate with accessible data.

In this section, we first examine the discrete-space models which revolutionized the measurement of agglomeration and then briefly discuss some of their derivative measurements. Next, in Section 2.2, we discuss the model-free discrete-space measurements, primarily as a segue into our treatment of the continuous-space agglomeration indices.

2.1. **Discrete Indices.** In the first paper to provide a satisfactory index of industry agglomeration, Ellison and Glaeser (1997) treated agglomeration as the combined effect of natural advantage and industry spillovers. In this model, N firms sequentially choose amongst M locations. An individual firm must choose whether to follow the prior firm's decision or to choose a location randomly by throwing a dart at a map.

From this model, Ellison and Glaeser (1997) derive the *EG-index* γ_ι^{EG}, given by[5]

$$\gamma_\iota^{EG} = \frac{\sum_{i=1}^M (s_i - x_i)^2 - \left(1 - \sum_{i=1}^M x_i^2\right) \sum_{j=1}^N z_j^2}{\left(1 - \sum_{i=1}^M x_i^2\right)\left(1 - \sum_{j=1}^N z_j^2\right)}$$

where s_i is the share of industry ι's employment in area i, x_i is the share of total employment in area i, and the $\{z_j\}$ are the sizes of the plants j of industry ι. Defining for an industry ι the *Gini index* $G_\iota = \sum_{i=1}^M (s_i - x_i)^2$ and the *Herfindahl index* $H_\iota = \sum_{j=1}^N z_j^2$, we obtain the more commonly-used expression

$$\gamma_\iota^{EG} = \frac{G_\iota - \left(1 - \sum_{i=1}^M x_i^2\right) H_\iota}{\left(1 - \sum_{i=1}^M x_i^2\right)(1 - H_\iota)} = \frac{G_\iota/(1 - \sum_{i=1}^M x_i^2) - H_\iota}{1 - H_\iota}.$$

The EG-index presents a number of advantages, which are discussed by Ellison and Glaeser (1997): First, it provides an unbiased estimate of agglomerative forces independent of the source

[4]These measurements may be slightly skewed by physical boundaries, as discussed in Marcon and Puech (2003).

[5]There is a small typographical error in the expression for γ_ι^{EG} given as equation (5) of Ellison and Glaeser (1997). In particular, there is a spurrious square of a $1 - \sum_{i=1}^M x_i^2$ term in the numerator of expanded expression; we have used the fixed expression given in Ellison, Glaeser, and Kerr (2007).

of such forces. This estimate is easily interpreted, as the probability that a firm choosing its location follows the prior firm rather than locating randomly. Further, γ_ι^{EG} is easily computed, as it only depends on spatial-unit level information about the plant distributions of the industry ι. The EG-index is comparable across industries with varying firm in size-distributions and controls for overall agglomeration trends. Also, the EG-index separates spatial aggregation from the measurement of agglomeration.[6]

Following the Ellison-Glaeser (1997) framework, Maurel and Sédillot (1999) proposed a modification of the EG-index and developed a new sequential model of firm location choice. First, they suggested the so-called *MS-index* of geographic concentration γ_ι^{MS},

$$\gamma_\iota^{MS} = \frac{(\sum_{i=1}^{M} s_i^2 - \sum_{i=1}^{M} x_i^2)/(1 - \sum_{i=1}^{M} x_i^2) - H_\iota}{1 - H_\iota}.$$

This index is based on the *weighted estimator*, which weights plant spillover measurements in each individual location i by the sizes of the plants in location i. It happens that $E(\gamma_\iota^{MS} - \gamma_\iota^{EG}) = 0$, so that Maurel and Sédillot (1999) may piggyback on Ellison and Glaeser (1997)'s well-definedness results and model, so that γ_ι^{MS} can also be seen as a measurement of the difference between firm distribution across industry ι and random chance. Unfortunately, this facet of the model is not developed further.

In the second half of their paper, Maurel and Sédillot (1999) envision a process in which firms first study the natural advantages and potential spillover benefits available and choose a region R in which to locate and then choose a more specific location $\ell \in R$ within each region, based on a region-level spillover model. From this model, they are able to compute weighted "second-stage" concentration measurements, which they find to be robust to the "first-stage" concentration measurements.[7]

The EG-index and MS-index give similar results on industry concentration.[8] This is not surprising, given the similarities in the structure of the two indices.

The Ellison-Glaeser (1997) framework, however, is not common to all indices; it is not even used in all of the discrete-space indicies. A newer breed of discrete-space indices has relied on carefully selected statistical tests at the expense of underlying models.

[6] With their index, Ellison and Glaeser (1997) measured the concentration levels of the 459 four-digit industries classified in the Census Bureau's 1987 SIC system. They found, at the state-level of spatial analysis, that over 97% of United States four-digit industries are agglomerated.

[7] That is, the agglomeration level rankings of the industries found with the second-stage concentration measurements are quite similar to those found using the first-stage measurements.

[8] Maurel and Sédillot (1999) compute both indices for French industries and note that the indices agree, in general, on the list of "most localized" industries. There is more variation between the two indices with respect to the least-localized industries, but this is not surprising when one considers the small variance of γ-values associated with low agglomeration levels. Devereux, Griffith, and Simpson (2003) compare the MS-indices for French and United Kingdom industries with the EG-indices of American firms and find results which support the basic findings of Maurel and Sédillot (1999).

First, Rysman and Greenstein (2005) developed a combinatoric test for agglomeration along the lines of local industry standardization, the *Multinomial Test for Agglomeration and Dispersion* (*MTAD*). Their index relies on the computation of

$$t(\mathbf{X}, \mathbf{n}, \mathbf{p}) = l(\mathbf{X}, \mathbf{n}, \mathbf{p}) - E\left[l(\mathbf{X}, \mathbf{n}, \mathbf{p})\right],$$

where they have defined

$$l(\mathbf{X}, \mathbf{n}, \mathbf{p}) = \frac{1}{M}\sum_{i=1}^{M} \ln\left[\binom{n_i}{x_i^1, \ldots, x_i^C}\right] + \ln\left[\prod_{k=1}^{C} p_k^{x_i^k}\right].$$

Here, M is the total number of markets where firms may locate, each populated by n_i firms. Firms choose between C options, available in each market, and the number of agents observed choosing option c in market i is written x_i^c, with $\mathbf{X} = (x_i^1, \ldots, x_i^C)$. The unconditional probability of choosing option c is denoted p_c, with $\mathbf{p} = (p_1, \ldots, p_C)$.

MTAD tests for agglomerative behavior of firms, under the presumption that firms are agglomerating if and only if they exhibit similar choice behaviors when colocated. The principal advantage of MTAD is that it effectively reports actual levels of dispersion; this feature is absent from the EG- and MS-indices, which only report whether dispersion exists.[9] While these theories are interesting and the dispersion model is certainly desirable, Rysman and Greenstein (2005) only test their index on a single industry, the United States Internet Service Providers.[10] This, combined with the fact that no model supporting MTAD is available, causes us to question the value of this index in the study of agglomeration.

More promising, perhaps, is the discrete-space, axiomatic model presented by Mori, Nishikimi, and Smith (2005). Their index is purely statistical, based on the *Kullback-Leibler divergence* (*D-index*) measurement.[11] This index is given, in theory, by

$$D\left(p_\iota \mid p_0\right) = \sum_{i=1}^{M} p_{\iota i} \ln\left(\frac{p_{\iota i}}{p_{0i}}\right),$$

where industries $\{\iota\}$ are located in M locations, with each industry ι having $N_{\iota i}$ firms (or factories) in region i. Here, also, the probability of a randomly sampled industry ι being located in region i is given by $p_{\iota i}$, and the reference distribution p_{0i} is the probability that a randomly sampled establishment will be in region i under a model of complete spatial dispersion. In practice, $p_{\iota i}$ is

[9]As Rysman and Greenstein (2005) note, this proposition follows immediately from an examination of the model which underlies the "dartboard" approach.

[10]Rysman and Greenstein (2005) further assert that "MTAD is easy to compute and interpret, and performs well in practice." We dispute this claim, on the basis that the empirical evidence is neither convincing nor easily generalizable to the types of data available for other industries.

[11]This measurement is introduced in Kullback and Liebler (1951).

not direcly observable, so that a natural sample estimate $\hat{p}_{\iota i}$ of $p_{\iota i}$ must be used,

$$\hat{p}_{\iota i} = \frac{N_{\iota i}}{\sum_{i=1}^{M} N_{\iota i}}; \quad D\left(\hat{p}_{\iota} \mid p_0\right) = \sum_{i=1}^{M} \hat{p}_{\iota i} \ln\left(\frac{\hat{p}_{\iota i}}{p_{0i}}\right) \approx D\left(p_{\iota} \mid p_0\right).$$

Mori, Nishikimi, and Smith (2005) propose to measure agglmeration through a comparison between the D-index for an industry and a reference model of complete spatial dispersion. This approach's validity stems from the fact that the D-index is a limiting form of the log-*likelihood ratio* for the hypothesis $p_i \equiv p_0$. This measurement is attractive because it is independent of sample size,[12] although it still suffers from the regular problems of the discrete-spatial region indices. The D-index can, further, be decomposed to give information about agglomeration across and among regional bundles, thereby giving a measurement similar to that obtained in the two-stage model of Maurel and Sédillot (1999).

The D-index values were computed for Japanese industries in Mori, Nishikimi, and Smith (2005) and were then compared to the associated Gini index values.[13] While the D-index and Gini index generally differ on a case-by-case basis, they are highly positively correlated for Japanese industries.[14] Nonetheless, the differences between their respective definitions of the spatial decentralization reference case make it especially difficult to make any realistic comparison between these two indices.[15]

All of these indices report similar conclusions, as we will see in Section 3.1. They benefit from (relatively) loose data requirements, and some also have clear interpretations derived from economic choice models. However, they all, as we have observed, suffer from the same deficiency: they aggregate firms across discrete spatial units. This assumption of discreteness, as noted by Mori, Nishikimi, and Smith (2005) and by Bertinelli and Decrop (2005), ignores geographic relationships between locations and is therefore a serious disadvantage.

2.2. **Continuous Indices.** We now turn, then, to indices which avoid the discrete-space assumption and work instead with continuous spatial models. These measurements were developed in

[12] Note that sample size does, however, play a role in the confidence bounds on the computed values of D.

[13] Mori, Nishikimi, and Smith (2005) were unable to obtain the establishmet-size data needed in order to run a comparison between the D-index and the EG- or MS-indices. Hence, they settled for a comparison of their index and the Gini index, which they argued is closely correlated with the EG-index. A quick look at Table 4 of Ellison and Glaeser (1997) shows that this is not a terrible assumption for highly localized industries but that more care may be needed when working with the more dispersed industries.

[14] It is suggested, however, that this is a side effect of the fact that Japanese industries are likely to have small employment shares whenever they are localized (see Mori, Nishikimi, and Smith (2005)). It would be interesting to do the same comparison for a country with dimensions more regular than Japan's.

[15] In their discussions, Mori, Nishikimi, and Smith (2005) suggest that one might create a few new indices $\hat{G}$, $\hat{D}$ to bridge the divide between the two methodologies, but their ideas seem fanciful. In particular, they suggest that one might obtain meaningful results by either computing the Gini index of their probability measures or the D-index of the the spatial organization variables of Ellison and Glaeser (1997); it is not clear how we would interpret such measurements.

direct response to the failures of the discrete-space models and often invoke complicated statistical techniques not common to the economic geography literature.

The first study of this sort was the distance-based measurement of Duranton and Overman (2005).[16] Using a measurement of bilateral distance density, Duranton and Overman (2005) derived a spatial agglomeration measurement which satisfies the core criteria (criteria (1)-(5) of Section 1) they argued an agglomeration index must satisfy. This measurement is also said to be derived from a model of spatial equilibrium and firm choice, although the details given are somewhat sketchy.[17]

In particular, they measured the distribution of geographical distances between pairs of firms in an industry and comp d these distributions with a hypothetical, random distribution of firms. They computed the K*-density*

$$\hat{K}(d) = \frac{1}{N(N-1)h} \sum_{j_1=1}^{N-1} \sum_{j_2=j_1+1}^{N} f\left(\frac{d - d_{j_1,j_2}}{h}\right),$$

which measures the density of bilateral distances d between firms. (Here, N is again the total number of firms, d_{j_1,j_2} is the distance between firms j_1, and j_2, h is a bandwith parameter and f is a Gaussian kernel.)

Interestingly, Duranton and Overman (2005) found strikingly different concentration results from those of Ellison and Glaeser (1997). As we mentioned above, Ellison and Glaeser (1997) found over 97% of United States industries to be agglomerated. By contrast, Duranton and Overman (2005) argued that only 43% of United Kingdom industries are agglomerated, while a striking 22% of United Kingdom industries are actually dispersed. They compare these values to self-computed values of the EG-index, which indicate that 94% of the industries in the United Kingdom are agglomerated in the Ellison and Glaeser (1997) framework. They also suggested that the cutoff Ellison and Glaeser (1997) proposed as an indicator of “high levels of agglomeration” is actually far too low.[18] Perhaps most importantly, Duranton and Overman (2005) illustrated the failure of the discrete approach by identifying two highly agglomerated United Kingdom industries which the EG-index fails to characterize as concentrated.[19]

Marcon and Puech (2003) used a framework similar to that of Duranton and Overman (2005), but used *Besag's* L*-function*, a normalized form of *Ripley's* K*-function* (which, in turn, differs

---

[16]While it is not the first study of this sort to be published, the earliest published paper applying a continuous model (Marcon and Puech (2003)) cites Duranton and Overman (2005)'s working paper, Duranton and Overman (2002).

[17]In an appendix, Duranton and Overman (2005) invoke Proposition 1 of Ellison and Glaeser (1997), without adequately justifying why the spatial distance model's constraints actually satisfy the conditions of Ellison and Glaeser (1997)'s framework. It is the opinion of the author that the result is misapplied.

[18]They gave further evidence for this proposition, using United Kingdom industries, which they then compared to Devereux, Griffith, and Simpson (2003).

[19]The United Kingdom's publishing and jewelry industries are both localized around London. The greater London area is divided into a large number of postal codes, so that the EG-index fails to identify the local concentration of these industries at the postal-code level.

slightly from Duranton and Overman (2005)'s own K-function) to compute their measurement of spatial localization. This framework lacks a supporting model, although it may be possible to adapt the model of Duranton and Overman (2005).

A nice comparison was given between the L-function results and the K-function results found by Duranton and Overman (2005). While Marcon and Puech (2003) made the somewhat silly argument that Duranton and Overman (2005) should not have attempted to build a tool from scratch when a suitable tool existed, they also gave an effective critique of the quantification failures of Duranton and Overman (2005)'s K-function.[20] Several advantages of Duranton and Overman (2005)'s K-function over the L-function were identified, as well.[21]

These two continuous measurements share one core strength: they are based on absolute distance measurements and hence are independent of spatial unit size choice. They are not prone to the spurious correlations which arise during the aggregation processes in the discrete models. However, both Duranton and Overman (2005)'s and Marcon and Puech (2003)'s studies required massive amounts of data—they needed the exact spatial addresses of every firm in their sample. These measurements were also computationally intensive to calculate. Further, while it is easy to test the statistical significance of these continous models using Monte Carlo methods, such methods make some of the results irreproducible.

Finally, we turn to the very promising working paper of Guillain and Le Gallo (2007), which combined discrete-space and continuous-space models of spatial agglomeration. Guillain and Le Gallo (2007) focused on the distinction between *clustering*, which they argued can be identified by the discrete-space measurements, and *agglomeration*, for which clustering is a necessary but not sufficient component.

Owing to deficiencies in their data, Guillain and Le Gallo (2007) were only able to compute the Gini index and the *Moran's I coefficient*[22] of a region, rather than any of the more advanced discrete-space statistics.[23] They then used exploratory spatial measures to look for spatial autocorrelation, which would be indicative, they claim, of actual agglomeration. For this task, Guillain and Le Gallo (2007) employed both *Moran scatterplots* and *Local Indicators of Spatial Associations* (*LISA statistics*). These indicators determine the levels of spatial clustering exhibited in the region surrounding an observation of concentration.

[20]The K-function does not effectively characterize levels of dispersion (see Marcon and Puech (2003)).

[21]The most pertinent is that the K-function can be modified to control for firm size and industrial concentration. Also, the K-function need not be corrected for edge-effects.

[22]Since this second statistic is difficult to explain and not especially relevant to our discussion, we omit the definition here. See Guillain and Le Gallo (2007) or Cliff and Ord (1981) for more information.

[23]These issues are not salient, however, across most data sets available; we expect that the methodology of Guillain and Le Gallo (2007) could be improved through the use of any of the advanced measurements discussed in Section 2.1 and that such improvements are realistic with the tools at hand.

The most important benefit of the exploratory spatial statistics is that they obviate the necessity of arbitrary cutoffs for agglomeration levels. In this way, Guillain and Le Gallo (2007) dealt with one of Duranton and Overman (2005)'s principal objections to Ellison and Glaeser (1997). The spatial statistics methods also allow for natural statistical significance assessments. Overall, these techniques applied by Guillain and Le Gallo (2007) are quite impressive and make great strides towards reconciling the discrete-space and continuous-space approaches. Indeed, the only significant deficiency we can find in their framework is the lack of a model.[24]

3. The Dynamics of Agglomeration

As we discussed in Section 1, the key purpose of the agglomeration measurement literature is to provide a consistent measurement of relative agglomeration levels. Such a measurement can then be used to understand the pathways through which agglomeration leads to growth.

Beyond the measure-proposing papers we discussed in Section 2, there have been numerous papers which computed one or more of the indices for different data sets. We survey some highlights of this work in Section 3.1, below, and discuss its relation to the rest of the literature. Then, in Section 3.2, we discuss some of the most interesting applications of the agglomeration indices to studies of agglomeration dynamics.

3.1. **Index Case Studies.** A fantastic amount of literature has been developed out of the computation of the EG-index. As we recall, Ellison and Glaeser (1997) computed the EG-index for United States industries. Barrios *et al.* (2004) computed the EG-indices of industries in Portugal and Ireland, with an eye towards understanding firm mobility dynamics.[25] Duranton and Overman (2005) computed the EG-index for industries in the United Kingdom, finding similar results to those of Ellison and Glaeser (1997). Bertinelli and Decrop (2005) computed the EG-indices of Belgian manufacturing industries, and compared the values they obtained to Moran I index values. Maré (2005) computed the EG-index values for industries in New Zeland. In general, these studies found large degrees of agglomeration, especially amongst traditional industries.

Maurel and Sédillot (1999) computed their MS-index's values for French industries. Devereux, Griffith, and Simpson (2003) computed the MS-indices of United Kingdom industries, finding that these index values were closely related to the EG-index values for United States industries and to the MS-indices for French insustries. These studies found evidence suggesting that high-tech industries are less agglomerated than traditional, low-tech industries; this trend matches the results found with the EG-index.

Only Mori, Nishikimi, and Smith (2005) computed agglomeration index values for Japan; indeed, this was the only Asian agglomeration index study of interest which we found. This is

[24]We also note that no mention of the computational complexity of these measures was given, so that we can only theorize about the difficulties Guillain and Le Gallo (2007) may have encountered in computing the measurements.
[25]This study will be examined in more depth in Section 3.2.

unfortunate, as the measurement techniques of Mori, Nishikimi, and Smith (2005) are highly independent from the other indices we have discussed. The lack of other Asian industry studies for comparison exacerbates the already great difficulty of evaluating this index's effectiveness.[26]

The continuous models have not been widely applied, as it is very difficult to obtain the data required. Indeed, Bertinelli and Decrop (2005) remarked that they would prefer to have used a continuous measurement for their study but were unable to obtain sufficiently detailed data.

Thus, the empirical literature on continuous measurements of agglomeration can only draw on the foundational papers' computations. Duranton and Overman (2005) computed their K-index for United Kingdom industries, at varying levels of detail. Both Marcon and Puech (2003) and Guillain and Le Gallo (2007) computed their indices for French industries, centering on the Paris area. While Guillain and Le Gallo (2007) do cite Marcon and Puech (2003), they do not undertake a comparison between the results. Our inspection reveals that the industry classifications in the two papers differ starkly. Further, Guillain and Le Gallo (2007) tend to disagree with Marcon and Puech (2003) along several important dimensions, including the relative agglomeration status of the textile industry.[27]

3.2. **Measuring Agglomeration Dynamics.** With the recent surge of agglomeration index literature, attention has also been given to the potential applications of agglomeration measurements. In general, these studies have relied on the discrete indices which give rise to clear models.

An early study in this vein, Audretsch and Feldman (1996), used the Gini index to identify links between spatial agglomeration in manufacturing industries and industry-specific characteristics. In particular, knowledge spillovers were found to be especially important in determining levels of industry agglomeration.

Ellison and Glaeser (2002) applied the EG-index to a study of the importance of natural advantage to industrial agglomeration. This study concluded that 20% of measured agglomeration levels arise through the natural advantage pathway; Ellison and Glaeser (2002) conjectured the actual explanatory power of natural advantage to be closer to 50%.[28]

In a comprehensive study, Rosenthal and Strange (2001) regressed the EG-index on a selection of important industry characteristics: knowledge spillovers, labor market pooling, input sharing,

[26]Recall that, as mentioned in Section 2.1, above, the D-index of Mori, Nishikimi, and Smith (2005) is generally incomparable to the other indices.

[27]Guillain and Le Gallo (2007) seem to suggest that some traditional industries in France are not as agglomerated as the other studies claim. This might be seen to cast doubt upon their computational methods, as it appears that most of the confusion is arising from the Gini index values found in their study.

[28]This study suffered from several deficiencies, notably the small number of advantage variables considered and the (seemingly unnecessary) restriction of the study to manufacturing industries. A more salient, methodological problem which is barely discussed by Ellison and Glaeser (2002) is the ignorance of physical spatial characteristics inherent in the discrete model; this distortion abstracts the notion of local natural features.

product shipping costs, and natural advantage. Their study attempted to control the spatial distortion, by working with a variety of spatial units. The results showed strong effects of all of these factors, with labor market pooling exhibiting the strongest correlation with agglomeration.[29]

Dumais, Ellison, and Glaeser (2002) sought to describe the dynamics of agglomeration. This study, working with United States manufacturing industry data and the EG-index, showed that new firms' location choices serve to decrease levels of agglomeration, whereas plant closures often reinforce agglomerative trends. This result was confirmed by Barrios *et al.* (2004), which conducted a similar EG-index study over Irish and Portuguese manufacturing industries.

Conducting a similar study, Holmes and Stevens (2002) showed that firms generally build large plants in areas of agglomeration and build smaller plants in less agglomerated regions. They developed their own framework, which was similar to that of Ellison and Glaeser (1997)'s dartboard model. Recently, Devereux, Griffith, and Simpson (2007) suggested an interesting proposition which seems to extend this result: pre-existing agglomerative structures affect firms' location decisions upon industry entrance. In particular, Devereux, Griffith, and Simpson (2007) obtained the result that firms are less responsive to government subsidies in areas having fewer pre-existing plants in their industry.

4. Conclusion

Having examined seven different indices of agglomeration, we are now in a position to draw conclusions about the most important directions for future study. Foremost, since application of the continuous indices has proven cumbersome, we know that none of the continuous indices are sufficient for our purposes, at present. However, the discrete indices also have their problems, as Duranton and Overman (2002)'s London example illustrates so well.

It is possible that the optimal solution to the agglomeration index problem will be a combination of measurements, after the style of Guillain and Le Gallo (2007). However, if this is the case, then we expect the key insights to come from clever, axiomatic, application of regional science. It is, unfortunately, not clear how to model such an index.

Thus, despite the difficulties, we suspect that the answer may lie in a clever modeling application of a single statistical tool. We do not believe this tool has been found yet.

References

Audretsch, David B, and Feldman, Maryann P, 1996. "R&D Spillovers and the Geography of Innovation and Production," *The American Economic Review* **86**(3): 630–640.

[29]Ellison, Glaeser, and Kerr (2007) conducted a similarly complete study of agglomeration dynamics through the prism of coagglomeration patterns, using a coagglomeration index based on the EG-agglomeration index (also first introduced in Ellison and Glaeser (1997)). They found similar results, except that they identified input-output dependencies as more important than labor market pooling to agglomeration levels.

Barrios, Salvador, Bertinelli, Luisito, Strobl, E, and Teixeira, Antonio-Carlos, 2004. "The dynamics of agglomeration: evidence from Ireland and Portugal," *Journal of Urban Economics* **57**: 170–188.

Bertinelli, Luisito, and Decrop, Jehan, 2005. "Geographical Agglomeration: Ellison and Glaeser's index Applied to the Case of Belgian manufacturing Industry," *Regional Studies* **39**(5): 567–583.

Ciccone, Antonio, and Hall, Robert E, 1996. "Productivity and the Density of Economic Activity," *American Economic Review* **86**(1): 54–70.

Cliff, A. D, and Ord, J. K, 1981. *Spatial Processes: Models and Applications*, London: Pion.

Devereux, Michael P, Griffith, Rachel, and Simpson, Helen, 2003. "The geographic distribution of production activity in the UK," *Regional Science and Urban Economics* **34**: 533–564.

Devereux, Michael P, Griffith, Rachel, and Simpson, Helen, 2007. "Firm location decisions, regional grants, and agglomeration externalities," *Journal of Public Economics* **91**: 413–435.

Duranton, Gilles and Overman, Henry G, 2002. "Testing for Localization Using Micro-Geographic Data," *The Review of Economic Studies* **72**(4): 1077–1106.

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

Ellison, Glen and Glaeser, Edward L, 1999. "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?," *The American Economic Review* **89**(2): 311–316.

Ellison, Glen, Glaeser, Edward L, and Kerr, William, 2007. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," NBER Working Paper 13068.

Fujita, Masahisa and Thisse, Jacques-François, 2002. *The Economics of Agglomeration*, Cambridge: Cambridge University Press.

Glaeser, Edward L, Kallal, Hedi D, Scheinkman, Jose A, and Shleifer, Andrei. "Growth in Cities," *Journal of Political Economy* **100**(6): 1126–1152.

Guillain, Rachel and Le Gallo, Julie, 2007. "Agglomeration and dispersion of economic activities in Paris and its surroundings : An exploratory spatial data analysis," Regional Economics Applications Laboratory Discussion Paper 06-T-10.

Holmes, Thomas J, and Stevens, John J, 2002. "Geographic concentration and establishment scale," *Review of Economics and Statistics* **84**(4): 682–690.

Krugman, Paul, 1991. "Increasing Returns and Economic Geography," *The Journal of Political Economy* **99**(3): 483–499.

Kullback, Solomon, and Liebler, Richard A, 1951. *Information Theory and Statistics*, New York: Wiley.

Marcon, Eric, and Puech, Florence, 2003. "Evaluating the geographic concentration of industries using distance-based methods," *Journal of Economic Geography* **3**(4): 409–428.

Maré, D. C, 2005. "Concentration, Specialisation and Agglomeration of firms in New Zealand," Motu Economic and Public Policy Research Working Paper 0512.

Marshall, Alfred. *Principles of Economics: An Introductory Volume*, 8th ed. London: Macmillan, 1920.

Maurel, Francoise and Sédillot, Béatrice, 1999. "A measure of the geographic concentration in French manufacturing industries," *Regional Science and Urban Economics* **29**: 575-604.

Mori, Tomoya, Nishikimi, Koji, and Smith, Tony E, 2005. "A Divergence Statistic for Industrial Localization," *The Review of Economics and Statistics* **87**(4): 635–651.

Quah, Danny, 2002."Spatial Agglomeration Dynamics," *American Economic Review* **92**(2), 247–252.

Rosenthal, Stuart S, and Strange, William C, 2001. "The Determinants of Agglomeration," *Journal of Urban Economics* **50**: 191–229.

Rysman, Marc, and Greenstein, Shane, 2005. "Testing for agglomeration and dispersion," *Economics Letters* **86**(3): 405–411.

HTTP://WWW.SCOTTKOM.COM
E-mail address: kominers@fas.harvard.edu