

This is an accepted manuscript of an article published by Springer in *The Annals of Regional Science* on 13 August 2014. The final publication is available at Springer via <http://dx.doi.org/10.1007/s00168-014-0627-z>

Please cite as:

Dubé, Jean, and Cédric Brunelle. 2014. "Dots to dots: A general methodology to build local indicators using spatial micro-data." *The Annals of Regional Science* 53 (1):245-272. doi: 10.1007/s00168-014-0627-z.

Dots to dots: A General Methodology to Build Local Indicators using Spatial Micro Data

Jean Dubé,¹ Professor,
Université du Québec à Rimouski (UQAR)
300, allée des Ursulines
Rimouski, Québec, Canada, G5L 3A1
jean_dube@uqar.qc.ca

Cédric Brunelle, Assistant Professor
Memorial University
230, Elizabeth Ave
St-John's, Newfoundland, Canada, A1B 3X9
cedric.brunelle@ucs.inrs.ca

¹ This research is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC).

Dots to dots: A General Methodology to Build Local Indicators using Spatial Micro Data

Empirical studies in regional science have so far largely relied on discrete conceptualizations of space and aggregated metrics which do not take into consideration spatial heterogeneity and variability at the micro level. This paper explores the use of these indicators when dealing with observations at the sub-regional level, based on micro data sets that impose the conceptualization of spatial interactions in a continuous and multidirectional space. We propose a general methodology to build local indicators for spatial micro data sets. Based on distance matrix and matrix calculation, some classical indices of specialization and diversity are extended to their local counterparts to explore the full spatial heterogeneity and variability of space. The methodology is applied to 9,839 establishments covering all economic sectors in the Lower St-Lawrence (LSL) region (Quebec, Canada). We find that the distribution of the local indicator varies significantly with distance, which suggests that the effects of specialization or diversity are not constant over space. Treating space as continuous may become of prime importance, given that more individual data sets are now available and the performance of microcomputers is still improving.

Key words:

Specialization and diversity indices; Local indices; Regional development; Spatial analysis.

Introduction

Theorists have long acknowledged the influence of local industrial structures on regional growth and the agglomeration of firms (Marshall, 1890; Jacobs, 1969). Industrial specialization (henceforth *specialization* for short) and regional diversity (*diversity*) have been among the most prolific and widely tested regional characteristics, giving rise to the development of a wide range of indicators and metrics over time. While empirical evidence have generally supported the positive impacts of both specialized clusters and diversified cities (Duranton and Puga 2000), studies have also frequently exhibited conflicting results which have made the specialization-diversity debate largely inconclusive in the literature (Beaudry and Schiffauerova 2009; Shuai 2013; Rosenthal and Strange 2001). One reason is that most analyses have so far relied on aggregated data and indicators which do not take into account spatial heterogeneity and variability at the micro level. Authors have argued that estimations can significantly be influenced by measurement and aggregation issues—indicators of specialization and diversity being particularly sensitive to small variations in their distributions (Beaudry and Schiffauerova 2009; Rosenthal and Strange 2004). As a consequence, studies tend to treat clusters as quasi-homogenous with no internal variability, and conceptualize space as *discrete* rather than *continuous*. As noted by Rosenthal and Strange (2004), “*relatively little of the empirical work on the scope of agglomeration economies has addressed the issues of establishment uniqueness and continuity*” (Rosenthal and Strange, 2004, p.5). Defining the actual scale and scope of spatial effects below aggregated areas thus represents an open empirical challenge, with potentially important implications for local development policies (van Soest et al., 2006).

Recently, a growing number of studies have raised the possibility of exploiting spatial micro data (see Aruzo-Carod et al., 2010). For instance, ... While these emerging approaches bring new explanations for regional dynamics, they also redefine the way researchers have to use and apply the traditional indicators, which involve the passage from a discrete to a continuous conceptualization of spatial relations. Despite its importance, this transition has so far remained largely unremarked in the literature. By proposing a new approach for integrating the variability of observations at the sub-regional level (i.e. individuals or firms), this paper aims to contribute to the methodological literature on spatial indicators in regional science. We present a general methodology to build local indicators for micro data sets that are continuous across space, and apply it on three widely used indicators in regional sciences: the Herfindahl index (H), the Shannon entropy index (E), and the location quotient (QL).

Building indices based on a continuous space may offer a significant contribution to the theoretical debate surrounding the effects of local spillovers and spatial externalities on individual establishments, clusters, or labour markets—topics that are central to establishing micro-economic foundations for regional economic development. It also provides a practical solution for measuring heterogeneity patterns of unique observations (individuals, firms, plants, or establishments) across space, while introducing spatial variability in the classical indices. In addition, it allows to test for the effects of spatial scales on the estimated outcomes.

Using micro spatial data, distance (weights) matrices, and matrix calculation, the methodology is applied on 9,839 individual establishments in the Lower St-Lawrence (LSL) region (Quebec, Canada). We provide evidence of the importance of taking advantage of spatial micro data before analysing and discussing properties of the index over space.

The paper is divided into four sections. In the first section, we present the main weaknesses associated with the use of indicators based on aggregated entities. We then formalize the mathematical construction of the indices, offering a short literature review on the use of specialization and diversity indices. The second section presents the general methodology for developing local indicators and exploring the possibility of varying the spatial scale using different distance thresholds. We then provide an empirical illustration of the potential applications in regional science, based on the analysis of business establishments data in the Lower-St-Lawrence (LSL) region (Quebec, Canada). The LSL represents a classical peripheral region where problems related to spatial aggregation can potentially influence the conclusion drawn from analysis using census tracking data, given that this aggregation consists of large geographic delimitations. The fourth section offers a brief discussion about the implication of developing local indicators for further research in regional sciences and economic geography.

1. Indices, Limits, and Scale

1.1 On the use of indices and fix spatial limits

Empirical studies in regional science, urban economics and economic geography largely rely on spatial data consisting of *a priori* defined geographic boundaries, such as municipalities, regions, provinces, states, and country. The literature stresses at least three problems related to discrete conceptualizations of space: (1) the ecological fallacy, (2) the modifiable areal unit problem (MAUP), and (3) the reduction of the total variance of the indices related to aggregations.

The *ecological fallacy* (Robinson, 1950) consists of generalizing the conclusions obtained from an analysis based on aggregated spatial units to micro units forming the spatial aggregation. Aggregated spatial units do not make it possible to figure out how individual units may react to their unique local environment. This is one of the most important limitations to the interpretation of the results based on a discrete conceptualization of space, and probably one of the most inherent critics formulated for such empirical analysis.

The *modifiable areal unit problem* (MAUP) (Openshaw, 1977; 1984) states that changing the geographical limits or spatial boundaries can influence the results obtained. If one wish to adopt a different definition of the geographical limits, then it is possible that this aggregation will result in different values of a given indicator and, consequently, in different conclusions. In other words, results may vary depending upon the definition of the geographical limits which made inferences difficult.

The *reduction of the variance* depend on the number of spatial entities considered (N) as well as on the heterogeneity of these spatial units, even if the aggregation is based on some clear criteria (Cliff et al., 1975).² The more heterogeneous the establishments are, the more variance there is, reflecting the wide difference among the individual, and smaller, units.³ Since those indices are usually introduced in a regression model as independent variables to test whether specialization or diversity measures may be critical explanatory factors of the dependent variable selected, the variability of the index can have a direct effect on the probable significance of the estimated coefficients. If the variation is quite low, it is possible that the statistical model will be incapable of identifying any statistical relation between these indices and the dependent variable, even though there may be one.

1.2 Some indices of specialization and diversity

There is a plethora of indices acting as a proxy to measure the specialization or diversity. For the sake of simplicity, only three frequently used indices are considered here: 1) Herfindahl (H), 2) Shannon entropy (E) and the 3) location quotient (LQ). However, the generalization to other indices is straightforward.

The Herfindahl index for a given region k , H_k , defines the relative specialization of the region. Formally, the Herfindahl index is simply defined as the sum of the squared proportion (in percentage) of jobs (or establishments), p_{sk} , (equation 1) calculated using the economic activities in economic sector s , x_{sk} , over the total economic activity in the region, $x_{.k}$ (equations 2 and 3).

$$H_k = \sum_{s=1}^S p_{sk}^2 \quad (1)$$

Where

$$p_{sk} = \frac{x_{sk}}{x_{.k}} \quad (2)$$

And

$$x_{.k} = \sum_{s=1}^S x_{sk} \quad (3)$$

To simplify, it can be postulated that $x_{sk} = e_{sk}$ if one uses the total employment, or $x_{sk} = f_{sk}$ if one uses the number of firms (establishments), where e_{sk} is the total employment of the region k in the economic productive sectors s and f_{sk} is the total number of establishments in region k , within the economic productive sectors s . In all

² This approach is also explained in detail by Deaton (1985) and Verbeek and Nijman (1992) for constructing individual units for pseudo panel approaches.

³ Of course, the variance of the index, based on its distribution, can also be related to the choice of the number of economic sectors, S .

cases, the sum of the individual proportions, p_{sk} , for all economic sectors in a given region k , is equal to 1 (equation 4).

$$\sum_{s=1}^S p_{sk} = 1 \quad (4)$$

On one hand, it is easy to show that H_k reaches its maximum value when only one proportion is equal to 1, while the others are set to zero. This situation corresponds to the case where a region k is only active in one economic sector. On the other hand, H_k reaches its minimum value when the p_{sk} are all equal (uniform), that is, when economic activity is uniformly distributed. In this case, the Herfindahl index is equal to $1/S$. Low values of the Herfindahl index correspond to economic diversity, while high values correspond to a concentration of economic activities.

Another popular index used to measure relative specialization is the entropy (Shannon) index in region k , E_k (equation 5).

$$E_k = \sum_{s=1}^S p_{sk} \log \left(\frac{1}{p_{sk}} \right) \quad (5)$$

Where p_{sk} has the same interpretation as before (equation 2).

On one hand, the index E_k reaches its minimum value when only one p_{sk} is equal to 1, while the others are fixed to 0. In this context, the E_k index is equal to 0. On the other hand, the index admits no theoretical maximum values. Thus, the interpretation of this index is inverse to the Herfindahl index: a high value suggests an important diversity, while a value of 0 suggests a concentration in only one activity.

Another measure used in a wide range of empirical applications is the location quotient, LQ_{sk} (equation 6). The location quotient measures the relative specialization of a region k in the economic productive sectors s , p_{sk} , as compared to the mean concentration of the economic productive sector s in the economy, $p_{s\cdot}$.

$$LQ_{sk} = \frac{p_{sk}}{p_{s\cdot}} \quad (6)$$

In the last expression, p_{sk} has the same definition as before (equation 2), while the mean concentration of the economic productive sector s in the whole economy is defined as follows (equation 7),

$$p_{s\cdot} = x_{s\cdot} / x_{\cdot\cdot} \quad (7)$$

where,

$$x_{s.} = \sum_{k=1}^K x_{sk} \quad (8)$$

$$x_{..} = \sum_{s=1}^S \sum_{k=1}^K x_{sk} \quad (9)$$

The location quotient takes a minimum value of zero, when economic activity in sector s is not present in region k , while it admits no theoretical upper limits.

It is possible to standardize the location quotient to make sure that the domain has a value varying between -1 and 1 (equation 10).

$$NLQ_{sk} = \frac{LQ_{sk} - 1}{LQ_{sk} + 1} \quad (10)$$

The location quotient allows complementing the Herfindahl and Shannon entropy indices since it expresses the specialization or diversity in different economic sectors as compared to the global scale.

All these indices are used to carry out empirical investigations to test whether or not the specialization or diversity of a spatially aggregated unit can display an influence on a given phenomenon under study.

1.3 On the utility and the use of the indices

Economists have traditionally used these indices as proxies for different types of spatial spillovers and agglomeration economies. These proxies are then used to explore the possible links between economic performance of a spatial units (i.e. industry, region, or establishment), and the characteristics of these units, including specialization and diversity. Empirical applications are often based on the introduction of these indices as independent variables in econometric models in order to verify whether these have a significant influence on a dependent variable that usually measures the level of regional development or performance.

The typical approach is to define specialization and diversity measures through the aggregation of industries and plants in politically defined regions—Metropolitan Statistical Areas (MSAs) or counties for the US; Nomenclature of Territorial Units for Statistics (NUTS) for European regions; or Census Agglomerations and Divisions in Canada. In fact, the vast majority of studies still implicitly conceptualize space and spatial relationships as being *discrete*, based on either regional or sub-regional administrative units, in accounting for spatial externalities at the establishment level.

There exists many empirical investigation of the effect of external economies, as well as many variations in the results. Melo et al. (2009) show that the study characteristics influence the magnitude of the estimates. The spatial aggregation of data into spatial areal units may also explain why there is such variability in the results (De Dominicis et al., 2013; Leahy et al., 2010). De Groot et al. (2009) show that the level of regional aggregation is relevant for the strength at which the agglomeration forces are operational.

The availability of micro data opens the way to new challenges and new possibilities (Marcon and Puerch, 2010; 2003). However, many approaches using spatial micro data are still applying the classical scheme to construct specialization or diversity indices (Kronenberg, 2013; Mota and Brandao, 2013; Nakamura, 2012; Kosfeldt et al., 2011; Martin et al., 2011) or aggregate individual observations into some spatial areal units (Table 1). Thus, all these applications are subject to the MAUP.

INSERT TABLE 1 HERE

Some solutions have been proposed to deal with the MAUP. Briant et al. (2010) suggest using boxes instead of geo-political limits for France. Their analysis suggests that using administrative limits or a square geographical distribution have not much influence on the results. Recently, this spatial aggregation has been implemented by Andersson et al. (2012) and Larsson (2014) to Swedish data. Their analysis shows that the qualitative conclusions are not sensible to the spatial scale, while the magnitude of the elasticity parameters appear to be a smaller magnitude with smaller boxes. However, even using such approach, based on boxes or hexagons geographic forms (Christaller, 1933, Lösch, 1941),⁴ makes the analysis subject to the MAUP: applying a simple translation of the grid may results in highly different aggregations (Figure 1).⁵

INSERT FIGURE 1 HERE

The majority of the studies which are based on individual spatial units such as firms, plants, or establishments, omit the introduction of specialization or diversity indices as a way to avoid the MAUP. Many of these studies are based on a descriptive approach using spatial function, such as the Ripley's K-function (Albert et al., 2011; Duranton and Overman, 2005; Marcon and Puerch, 2003), point pattern analysis (Kosfeldt et al., 2011) or spatial autocorrelation statistics (De Dominicis et al., 2013). Other studies have exploited the potential of the micro data to study location and relocation decision, growth, productivity, survival pattern, clusters, subcontracting, exporting decision, and so on (Table 2). However, even these applications do not account for possible specialization or diversity through the usual indices.

INSERT TABLE 2 HERE

To our knowledge, there are few studies incorporating individual specialization and diversity indices in empirical analyses as independent variables. Li et al. (2012) propose a

⁴ Assuming that the optimal radius of influence is previously known.

⁵ Of course, the difference may be inversely related to the size of the boxes or hexagons.

simple extension of the location quotient based on the proposition of Holmes and Stevens (2002) by subtracting, for each point, its own employment level. By doing so, each point has a unique value of the location quotient reflecting the regional specialization excluding the given point. Ranski (2011) has proposed a more general measure based on a distance buffer allowing to retrieve, for each firm, a particular value of the location quotient that takes into account the spatial distribution of individual firms, plants, or establishments.

This paper proposes extending the specialization and diversity indices to local consideration by proposing a formal and general methodology to build such spatial indicators. The originality of the methodology lies in its simplicity, its rapidity of calculation, and the possibility of using the classical indices in empirical works based on individual spatial units while avoiding any spatial aggregation. Moreover, it allows taking space into account in the creation of specialization and diversity indices, which is usually not the case (Duranton and Overman, 2005).

2. Methodology

Building different types of indices using spatial micro-data may already represent a new challenge since accounting for the specialization or diversity of individual establishments (or points) involves accounting for a continuous and multidirectional space. To operationalize the indices in a continuous and multidirectional space (equations 1, 5 and 6), the methodological challenge consists of proposing a general definition of the neighbours, or the subscript k , for each individual unit. This can be conceptualized using spatial and spatio-temporal distance matrices (Dubé and Legros, 2013).

Assuming that observations consist of individual establishments denoted by $i, j = 1, 2, \dots, N$, and that the exact location of each establishment i is given by its geographical coordinates X_i, Y_i (Figure 2), it is possible to calculate the Euclidian distance separating all the establishments by using the usual Pythagorean theorem (equation 11).⁶

INSERT FIGURE 2 HERE

$$d_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \quad \forall i, j = 1, 2, \dots, N \quad (11)$$

The distances between all observations can then be synthesized through a general distance matrix, \mathbf{D} (equation 12), of dimension $(N \times N)$. A general line i expresses the distance among establishment i and all the others establishments. Since the distance separating an observation from itself is zero, the distance matrix contains elements equal to 0 on the principal diagonal.

⁶ Of course, the definition of the distance can be changed to implement any other distance criteria (Manhattan, network, etc.) using the general distance definition given by:

$$d_{ij} = \sqrt[p]{|X_i - X_j|^p + |Y_i - Y_j|^p} \quad \forall i, j = 1, 2, \dots, N$$

When $p = 2$, we obtain the usual Euclidian distance, while $p = 1$ gives the Manhattan distance. Moreover, other distances such as network distances can also be consider.

$$\mathbf{D} = \begin{bmatrix} 0 & d_{12} & \cdots & d_{1i} & d_{1j} & \cdots & d_{1N} \\ d_{21} & 0 & \cdots & d_{2i} & d_{2j} & \cdots & d_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ d_{i1} & d_{i2} & \cdots & 0 & d_{ij} & \cdots & d_{iN} \\ d_{j1} & d_{j2} & \cdots & d_{ji} & 0 & \cdots & d_{jN} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{Ni} & d_{Nj} & \cdots & 0 \end{bmatrix} \quad (12)$$

Using this general distance matrix, it is possible to build a connectivity matrix, \mathbf{C}_s , of dimension $(N \times N)$, that expresses a constraint version of the distance matrix based on a threshold distance, noted \bar{d} . This connectivity matrix can also be extended to each economic sector s (equation 13).⁷

$$c_{sij} = \begin{cases} x_{sj} & \text{if } d_{ij} \leq \bar{d} \text{ for } s \quad \forall i \neq j \\ 0 & \text{if } d_{ij} > \bar{d} \end{cases} \quad (13)$$

The critical distance cut-off value, \bar{d} , enables to vary the radius of influence and, in consequence, obtain unique values of the indices at each different spatial scale. By varying the value of the distance cut-off value, it is possible to generate a bunch of indices reflecting the specialization or diversity within a precise scale. Thus, the general \mathbf{C}_s matrix (equation 14) expresses, on a given line i , the spatial connectivity between an establishment i and the other establishments j , within the economic sector s .

$$\mathbf{C}_s = \begin{bmatrix} 0 & c_{s12} & \cdots & c_{s1i} & c_{s1j} & \cdots & c_{s1N} \\ c_{s21} & 0 & \cdots & c_{s2i} & c_{s2j} & \cdots & c_{s2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ c_{si1} & c_{si2} & \cdots & 0 & c_{sij} & \cdots & c_{siN} \\ c_{sj1} & c_{sj2} & \cdots & c_{sji} & 0 & \cdots & c_{sjN} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ c_{sN1} & c_{sN2} & \cdots & c_{sNi} & c_{sNj} & \cdots & 0 \end{bmatrix} \quad (14)$$

This mathematical expression offers a compact way to describe what may also be computed using a geographical information system (GIS) through a buffer specification varying the distance cut-off value (Figure 3). Yet, synthesizing and operationalizing these calculations through matrix calculations makes the computation more efficient, especially when using large databases.⁸

INSERT FIGURE 3 HERE

⁷ The connectivity matrix can be generalized to consider the inverse distance reflecting the first law of geography of Tobler (1970). Such transformation imputes a larger value to observations that are spatially closer.

⁸ The matrix notation accelerates the calculation time.

By noting $\mathbf{1}_N$ a vector of unity (1) of dimension $(N \times 1)$, it is possible to use the connectivity matrix \mathbf{C}_s and matrix operators to generate a set of vectors that can be used to calculate the different indices at the local scale. The first step is to measure the total economic activity in sector s surrounding each individual (micro) establishment (equation 15). This information is stocked in the vector \mathbf{x}_s of dimension $(N \times 1)$.

$$\begin{matrix} \mathbf{x}_s \\ (N \times 1) \end{matrix} = \begin{matrix} \mathbf{C}_s \\ (N \times N) \end{matrix} \times \begin{matrix} \mathbf{1}_N \\ (N \times 1) \end{matrix} \quad (15)$$

The total appearing in the first line of the vector \mathbf{x}_s represents the total number of establishments (if $x_{sj} = f_{sj} = 1$) or total number of jobs (if $x_{sj} = e_{sj}$) in the economic sector s within a distance of \bar{d} from the establishment 1, and so on.

Of course, the S individual vectors of dimension $(N \times 1)$ could be synthesized in a global matrix \mathbf{x} of dimension $(N \times S)$ ⁹. It is easy to see that the total number of establishments or jobs within a particular distance \bar{d} is given by the sum of the vectors \mathbf{x}_s over the S dimensions (equation 16).¹⁰

$$\begin{matrix} \mathbf{x} \\ (N \times 1) \end{matrix} = \begin{matrix} \mathbf{x}_1 \\ (N \times 1) \end{matrix} + \begin{matrix} \mathbf{x}_2 \\ (N \times 1) \end{matrix} + \dots + \begin{matrix} \mathbf{x}_S \\ (N \times 1) \end{matrix} \quad (16)$$

Ultimately, this matrix operation permits the retrieval of the vector of proportion \mathbf{p}_s in an economic sector s for a given establishment, or points (equation 17).

$$\begin{matrix} \mathbf{p}_s \\ (N \times 1) \end{matrix} = \begin{matrix} \mathbf{x}_s \\ (N \times 1) \end{matrix} \oslash \begin{matrix} \mathbf{x} \\ (N \times 1) \end{matrix} \quad (17)$$

Where \oslash is the Hadamard operator designating a term-by-term division.

As usual, the sum of the vectors \mathbf{p}_s over the s dimensions returns a vector of unity (equation 18).

$$\begin{matrix} \mathbf{p} \\ (N \times 1) \end{matrix} = \begin{matrix} \mathbf{p}_1 \\ (N \times 1) \end{matrix} + \begin{matrix} \mathbf{p}_2 \\ (N \times 1) \end{matrix} + \dots + \begin{matrix} \mathbf{p}_S \\ (N \times 1) \end{matrix} = \begin{matrix} \mathbf{1}_N \\ (N \times 1) \end{matrix} \quad (18)$$

Thus, the usual indices can easily be calculated at the individual level, given that the location of the individual points (or establishments) is known. Usual statistical models can thus be developed for individual spatial units.

The specification allows varying the cut-off distance parameter, \bar{d} , hence providing different local spatial indices for different values of \bar{d} . This general methodology can thus

⁹ Thus, we then have: $\begin{matrix} \mathbf{X}_s \\ (N \times S) \end{matrix} = \left[\begin{matrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_S \\ (N \times 1) & (N \times 1) & \dots & (N \times 1) \end{matrix} \right]$.

¹⁰ Using the definition of the global matrix, \mathbf{X}_s , the total of number of jobs or firms can be obtained with a simple matrix calculation: $\begin{matrix} \mathbf{x} \\ (N \times 1) \end{matrix} = \begin{matrix} \mathbf{X}_s \\ (N \times S) \end{matrix} \times \begin{matrix} \mathbf{1}_S \\ (S \times 1) \end{matrix}$, where $\mathbf{1}_S$ is a vector of one of dimension $(S \times 1)$.

explicitly deal with the impact of varying the distance cut-off value in empirical investigations. Moreover, using two different cut-off distances allows retrieving the local version of the location quotient, making it possible for the reference to vary spatially instead of being equal for all observations. This procedure may thus help identifying locally specialized or diversified economic activities.

The flexibility of the approach provides a way to calculate a set of critical distance cut-off values and compare the distribution of the indices according to this choice. It also allows testing the impact of defining a critical cut-off distance on the significance of the indices in statistical models. In specifying an econometric model, one can use different indices based on different value of \bar{d} and investigate, for example, the stability and amplitude of the coefficients as well as their significance. The choice of a particular critical distance can also be based on a statistical criterion that identifies the value of \bar{d} in the construction of the index for which the coefficient is the strongest and most significant, or one which minimizes any information criteria, such as Akaike or Schwartz. These possibilities are not formally undergone in the paper, but can easily be extended to particular empirical applications to formally test for possible aggregation problem (MAUP).

In other words, the use of a continuous spatial structure provides an indirect way, through the choice of \bar{d} , to take into consideration differences that may be relative to the scale of aggregation. It is then possible to calculate the indices at each given point across space and test the stability of the obtained results.

3. Empirical Application

3.1 Data

The following empirical application is based on an exhaustive list of individual establishments in the Lower St-Lawrence (LSL) region in Quebec, Canada (Brunelle and Dubé, 2013). The region is characterized by its remoteness from large metropolitan areas, low population densities, and an unequal distribution of establishments across space (Figure 5). The three main urban centres in the LSL are identified by light red shading on the map: Rimouski, Rivière-du-Loup and Matane, with populations, respectively, of 50,000, 20,000 and 15,000 inhabitants. Each urban centre has respectively 24,223, 14,766 and 7,821 jobs (Table 3) and 2,163, 1,296 and 777 establishments. The three main urban areas thus accommodate about 52% of the jobs and 43% (Table 4) of the total number of establishments. This contrasts with the region's size, which extends across twenty-two thousand square kilometres (22,404), distributed along 320 kilometres along the St-Lawrence River, making population densities relatively low.

INSERT TABLES 3 & 4 HERE

A consequence is the wide dispersion of establishments and activities across space (Figure 4). The rural character of the region can be seen in the high proportion of employment related to the primary economic sectors (Table 3) and the high number of

small establishments and the low number of large establishments (Table 4). In 2006, the database included a total of 9,839 establishments for a total of 88,939 jobs.

INSERT FIGURE 4 HERE

The database relies on the List of Industries and Trade (LIT) from the Quebec Ministry of Employment and Social Security (MESS). Spatial observations (establishments) are made available to their exact address, which allows geocoding and retrieving the exact geographical coordinates. The number of jobs is provided for each establishment, corresponding to the number of employees during the most active period of the year. The database reports the principal economic activity of the establishment by using the North American Industrial Classification System (NAICS) at the 6-digit level.

3.2 Descriptive Statistics of the Indices

The local indices of specialization and diversity are calculated for different cut-off values varying between 500 m. to 10,000 m. The value of \bar{d} increased by 500 m. at each step, procuring a full set of 20 local indices for each point. The development of the methodology using matrix expressions and calculations contributes to improving the speed of calculation. For example, calculating all the local indices for almost 10,000 observations for one distance cut-off value takes only 12 to 15 seconds.¹¹

As one should expect, the total variance of the indices is higher for lower cut-off distances, while showing a less concentrated distribution (Figures 6, 7 and 8).¹² This distribution clearly shows the impact on the variability of the indices according to the choice of the spatial scale. However, the gap becomes less pronounced when using establishments instead of total employment. This is linked to the fact that establishments are themselves of different sizes and are heterogeneously distributed over space. In contrast, the location of an establishment (a point) has an equal value for each observation (taking a value of 1), driving the distribution to be uniform in amplitude since providing less variability, while being not spatially random.

INSERT FIGURES 6, 7 & 8 HERE

Based on the total employment, the descriptive statistics for the Herfindahl and the entropy indices show that the variability of the values taken by the individual points is important (Table 5). For low values of the distance cut-off parameter, the standard error is higher for both indices, showing that the spatial structure can be quite heterogeneous, which can be seen through the individual local indices. This can also be observed through the increase and decrease of the mean value of the indices

INSERT TABLE 5 HERE

¹¹ Using the Mata platform on Stata software.

¹² The indices have been calculated for each cut-off distance value varying between 500 m. and 10,000 m. at each 500m. For simplicity sake and parsimony of the presentation, the results may available from the authors upon request.

One interesting feature of the distribution of the indices is that, in the actual case, over a critical distance cut-off of 5,000 metres, the principal descriptive statistics (mean and standard error) become quite stable for all specifications of the distance cut-off values. This is not surprising. If all observations were selected, only the individual characteristics of the firm (point) itself are not accounted for in the calculation of the local indices, returning more homogenous indices and considerably lowering the variability. Thus, at that level of aggregation, the only source of variability lies on the structure of the establishments themselves and not on the spatial distribution of the establishments.

The same conclusion can be drawn for the location quotient, even if some particularities exist in given economic sectors. With no surprise, the distribution of the indices shows that raising the distance cut-off value \bar{d} increases the number of establishments with higher specializations (value of QL_{sk} over 1), and reduces the variance of the distribution by lowering the possible values taken by the local location quotient. This supports, once again, a decrease in the variance of the indices when spatial aggregation is high, which is not without any consequences on possible statistical analyses and potential conclusions relative to the spillover effects of clustering. Thus, an inquiry on the size and scope of spatial specialization or diversity may be unable to identify beyond a given distance cut-off value.

In the end, the descriptive statistics of the local indices suggest that limiting the analysis to predefined geographical delimitations such as regions, cities or neighbourhoods can be an important limitation, especially when working with rural regions, since the geographical limits are actually quite large. Thus, the heterogeneity of the local proximity is usually drowned in the spatial aggregation, making it problematic to perform a precise analysis depending on the scale of the aggregation. These general local indices may be quite important when studying local phenomenon while trying to identify the effect of specialization or diversity at different scales. This may also be a major point if analyses are built upon a remote area such as the Lower St-Lawrence region.

4. Conclusion

In this paper, a general methodology to integrate local specifications of specialization and diversity indicators is developed for individual spatial (micro) data. The approach specifically allows to empirically apply conceptualizations of spatial relations as continuous rather than discrete. The general methodology can be used to deal with some drawbacks associated with traditional indicators based on aggregated spatial units, such as the ecological fallacy, the modifiable areal unit problem (MAUP), and the reduction of heterogeneity (and variance). Despite the fact that it remains practically impossible to entirely solve the MAUP, the proposed general methodology clearly addresses the possibility of identifying and estimating potential aggregation biases occurring by using different cut-off values. In other words, without explicitly solving the MAUP, the proposed methodology has the benefit of providing a clear way of checking for possible aggregation effects.

We illustrated potential applications of our methodology through an empirical analysis based on 9, 839 individual firms (establishments) in the Lower St-Lawrence region

(Quebec, Canada). This exercise has allowed building local indices and testing for their variability according to the choice of the distance cut-off values. The results demonstrate that the variability of a given index changes depending on the choice of the spatial scale. This observation is of primal importance if researchers want to make statistical inferences regarding the effects of specialization or diversity on a dependent variable, given that the choice of the spatial aggregation may likely influence the empirical results. This suggests that the significance of some parameters in econometric applications may be dependent on the amplitude of the variance of the independent variables. An independent predictor with a lower variability may lead to fewer chances of observing significant relations.

What is also noteworthy with this methodology is the fact that it becomes possible to determine the effects of changing the distance cut-off value. The consideration of different cut-off distances reveals also some other interesting results: the specialization and diversity indices may not lead to significant results at all distances. Without necessarily being the outcome of the MAUP, this could also identify some key conclusion, such as the fact that the effects of specialization or diversity are not constant over space. For instance, we may find that testing for the presence of spatial concentration or diversity at a given distance is significant, while becoming non-significant beyond a given threshold, which is consistent with the literature on spatial spillover.

It follows from the above that this methodology can have several implications for the study of regional dynamics, particularly in accounting for the effects of clustering across space. For example, it may allow the evaluation of the effect of local concentration depending on the economic sectors and identify possible spatial spillovers. Although these effects are well documented in the literature, there is no consensus on to what the spatial extent of these effects may be. By varying the distance cut-off value, it could prove possible to identify a critical distance over which the coefficient related to the indices measuring scale economies is not significant.

The methodology thus allows for a detailed investigation of the effects of spatial location and distribution. Future research could focus on improving the generalization of these particular applications on other micro-data sets and test for the spatial influence of specialization and diversity as well as other types of indicators. What would be interesting, for instance, is to determine whether the variation in the construction of the local indices has an impact on the significance of the parameter related to econometric modeling.

We think that this general methodology will be helpful for researchers who intend to work with individual spatial units and hope that the general notation will make it easy to implement on any software and exploit the full potential of the spatial micro data.

References

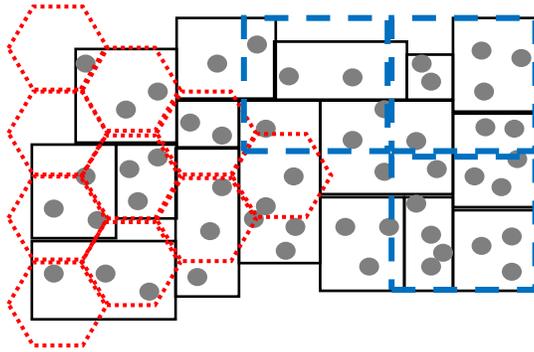
- Alama-Sabater, L., Artal-Tur, A. and Navarro-Azorin, J.M. (2011). Industrial Location, Spatial Discrete Choice Models and the Need to Account for Neighbourhood Effects, *Annals of Regional Science*, 47: 393-418.
- Albert, J.M, Casanova, M.R. and Orts, V. (2011). Spatial Location Patterns of Spanish Manufacturing Firms, *Papers in Regional Science*, 91(1): 107-136.
- Altomonte, C. and Colantone, I. (2008). Firm Heterogeneity and Endogenous Regional Disparities, *Journal of Economic Geography*, 8(2008): 779-810.
- Andersson, M. and Lööf, H. (2011). Agglomeration and Productivity: Evidence from Firm-Level Data, *Annals of Regional Science*, 46: 601-620.
- Andersson, M., Klaesson, J. and Larsson, J.P. (2012). *How Local are Spatial Density Externalities? Evidence from Square Grid Data*, CIRCLE Working Paper no. 2012/10, Lund University.
- Aruzo-Carod, J.-M. and Viladecans-Marsal, E. (2009). Industrial Location at the Intra-Metropolitan Level: The Role of Agglomeration Economies, *Regional Studies*, 43(4): 545-558.
- Aruzo-Carod, J.-M., Liviano-Solis, D. and Manjon-Antolin, M. (2010). Empirical Studies in Industrial Location: An Assessment of their Methods and Results, *Journal of Regional Science*, 50(3): 685-711.
- Baldwin, J.R., Brown, W.M. and Gu, W. (2012). Geographic Market Access and the Effects of Trade on Length of Production Run, Product Diversity and Plant Scale of Canadian Manufacturing Plants, 1974-1999, *Journal of Economic Geography*, 12(2012): 455-484.
- Baldwin, J.R., Brown, W.M. and Rigby, D.L. (2010). Agglomeration Economies: Microdata Panel Estimates From Canadian Manufacturing, *Journal of Regional Science*, 50(5): 915-934.
- Barlet, M., Briant, A. and Crusson, L. (2013). Location Patterns of Service Industries in France: A Distance-based Approach, *Regional Science & Urban Economics*, 43(2012): 338-351.
- Beaudry, C. and A. Schifffauerova. 2009. Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy* 38:318-337.
- Boschma, R. Eriksson, R. and Lindgren, U. (2009). How does Labour Mobility Affect the Performance of Plants? The Importance of Relatedness and Geographical Proximity, *Journal of Economic Geography*, 9(2009): 169-190.
- Briant, A., Combes, P.-P. and Lafourcade, M. (2010). Dots to Boxes: Do the Size and Shape of Spatial Units Jeopardize Economic Geogaphy Estimations?, *Journal of Urban Economics*, 67(2010): 287-302.
- Brunelle, C. and Dubé, J. (2013). *Does Proximity Increase the Survival of Firms in Remote Areas? An Analysis of Plant Closures in Quebec's Lower St. Lawrence Regions, 2006-2011*, Swedish Entrepreneurship Forum Working Paper, 2013:26, 45 p.
- Capasso, M., Cusmano, L. and Morrison, A. (2013). The Determinants of Outsourcing and Offshoring Strategies in Industrial Districts: Evidence from Italy, *Regional Studies*, 47(4): 465-479.
- Christaller, W. (1933). *Die Zentralen Orte in Studdeutschland*, Ilena University.
- Cliff, A.D., Haggett, P., Ord, J.K., Bassett, K.A. and Davies, R.B. (1975). *Elements of Spatial Structure*, Cambridge University Press, Cambridge.

- Daunfeldt, S.-O., Elert, N. and Rudholm, N. (2013). Start-ups and Firm in-migration: Evidence from the Swedish Wholesale Industry, *Annals of Regional Science*, 51: 479-494.
- De Dominicis, L., Arbia, G. and De Groot, H.L.F. (2013). Concentration of Manufacturing and Service Sector Activities in Italy: Accounting for Spatial Dependence and Firm Size Distribution, *Regional Studies*, 47(3): 405-418.
- De Vann, M, Boschma, R. and Frenken, K. (2013). Clustering and Firm Performance in Project-Based Industries: The case of the Global Video Game Industry, 1972-2007, *Journal of Economic Geography*, 13(2013): 965-991.
- Demirel, P. and Mazzucato, M. (2010). The Evolution of Firm Growth Dynamics in the US Pharmaceutical Industry, *Regional Studies*, 44(8): 1053-1066.
- De Groot, H.L.F., Smit, M.J. and Poot, J. (2009). Agglomeration Externalities, Innovation and Regional Growth: Theoretical Perspectives and Meta-Analysis, in Capello, R. and Nijkamp, P. (eds). *Handbook of Regional Growth and Development Theories*, Edward Elgar: 256-281.
- Deaton, A. (1985). Panel Data from Time Series of Cross Sections, *Journal of Econometrics*, 30(1-2):109-126.
- Dubé, J. and Legros, D. (2013). Dealing with Spatial Data Pooled over Time in Statistical Models, *Letters in Spatial and Resource Sciences*, 6(1): 1-18.
- Duranton, G. and Overman, H.G. (2005). *Testing for Localization using Micro-geographic Data*, LSE Research Online, <http://eprints.lse.ac.uk/archive/00000581>
- Duranton, G. and D. Puga. 2000. Diversity and specialisation in cities: Why, where and when does it matter? *Urban Studies* 37:533-555.
- Ellison, G. and E. L. Glaeser (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach, *Journal of Political Economy*, 105(5): 889-927.
- Eriksson, R. and Lindgren, U. (2009). Localized Mobility Clusters: Impacts of Labour Market Externalities on Firm Performance, *Journal of Economic Geography*, 9(2009): 33-53.
- Farole, T. and Winkler, D. (2014). Firm Location and the Determinants of Exporting in Low- and Middle-income Countries, *Journal of Economic Geography*, 14(2014): 395-420.
- Figueiredo, O., Guimaraes, P. and Woodward, D. (2010). Vertical Disintegration in Marshallian Industrial Districts, *Regional Science & Urban Economics*, 40(2010): 73-78.
- Figueiredo, O., Guimaraes, P. and Woodward, D. (2009). Localization Economies and Establishment Size: was Marshall Right After all?, *Journal of Economic Geography*, 9(2009): 853-868.
- Fontagné, L., Koenig, P., Mayneris, F. and Poncet, S. (2013). Cluster Policies and Firm Selection: Evidence from France, *Journal of Regional Science*, 53(5): 897-922.
- Giunta, A., Nifo, A. and Scalera, D. (2012). Subcontracting in Italian Industry: Labour Division, Firm Growth and the North-South Divide, *Regional Studies*, 46(8): 1067-1083.
- Graham, D.J. and Kim, H.Y. (2008). An Empirical Analytical Framework for Agglomeration Economies, *Annals of Regional Science*, 42: 267-289.
- Heebels, B. and Boschma, R. (2011). Performing in Dutch Book Publishing 1880-2008: The Importance of Entrepreneurial Experience and the Amsterdam Cluster, *Journal of Economic Geography*, 11(2011): 1007-1029.

- Holl, A. (2012). Market Potential and Firm-level Productivity in Spain, *Journal of Economic Geography*, 12(2012): 1191-1215.
- Holl, A., Pardo, R. and Rama, R. (2010). Just-in-Time Manufacturing Systems, Subcontracting and Geographic Proximity, *Regional Studies*, 44(5): 519-533.
- Holmes, T.J. and Stevens, J.J. (2002). Geographic Concentration and Establishment Scale, *The Reviews of Economics and Statistics*, 84(4): 682-690.
- Ibrahim, S.E., Fallah, M.H. and Reilly, R.R. (2009). Localized Sources of Knowledge and the Effect of Knowledge Spillovers: An Empirical Study of Inventors in the Telecommunications Industry, *Journal of Economic Geography*, 9(2009): 405-431.
- Jacobs, J. 1969. *The economy of cities*. Random House.
- Koenig, P. (2009). Agglomeration and the Export Decisions on French Firms, *Journal of Urban Economics*, 66(2009): 186-195.
- Kosfeld, R., Eckey, H.-F. and Lauridsen, J. (2011). Spatial Point Pattern Analysis and Industry Concentration, *Annals of Regional Science*, 47: 311-328.
- Kronenberg, K. (2013). Firm Relocations in the Netherlands: Why do Firms Move, and Where do They go?, *Papers in Regional Science*, 92(4): 661-713.
- Krugman, P. (1991). Increasing Returns and Economic Geography, *Journal of Political Economy*, 99: 493-499.
- Larsson, J.P. (2014). The Neighborhood or the Region? Reassessing the Density-Wage Relationship using Geocoded Data, *Annals of Regional Science*, online.
- Leahy, A., Palangkaraya, A. and Yong, J. (2010). Geographical Agglomeration in Australian Manufacturing, *Regional Studies*, 44(3): 299-314.
- Lee, Y. (2008). Geographic Redistribution of US Manufacturing and the Role of State Development Policy, *Journal of Urban Economics*, 64(2008): 436-450.
- Li, D., Lu, Y. and Wu, M. (2012). Industrial Agglomeration and Firm Size: Evidence from China, *Regional Science & Urban Economics*, 42(2012): 135-143.
- Lopez, R.A. and Sudekum, J. (2009). Vertical Industry Relations, Spillovers, and Productivity: Evidence from Chilean Plants, *Journal of Regional Science*, 49(4): 721-747.
- Lösch, A. (1940). *Die raumliche Ordnung der Wirtschaft*. Jena: Fisher.
- Marcon, E. and Puech, F. (2010). Measures of the Geographical Concentration of Industries: Improving Distance-based Methods, *Journal of Economic Geography*, 10(2010): 745-762.
- Marcon, E. and Puech, F. (2003). Evaluating the Geographic Concentration of Industries using Distance-based Methods, *Journal of Economic Geography*, 3(2003): 409-428.
- Marshall, A. 1890. *Principles of Economics*. London: Macmillan.
- Martin, P., Mayer, T. and Mayneris, F. (2011). Spatial Concentration and Plant-level Productivity in France, *Journal of Urban Economics*, 69(2011): 182-195.
- Melo, P.C., Graham, D.J. and Noland, R.B. (2009). A Meta-analysis of Estimates of Urban Agglomeration Economies, *Regional Science & Urban Economics*, 39(2009): 332-342.
- Mota, I. and Brandao, A. (2013). The Determinants of Location Choice: Single Plants versus Multi-plants, *Papers in Regional Science*, 92(1): 31-49.
- Nakamura, R. (2012). Contributions of Local Agglomeration to Productivity: Stochastic Frontier Estimations from Japanese Manufacturing Firm Data, *Papers in Regional Science*, 91(3): 569-597.

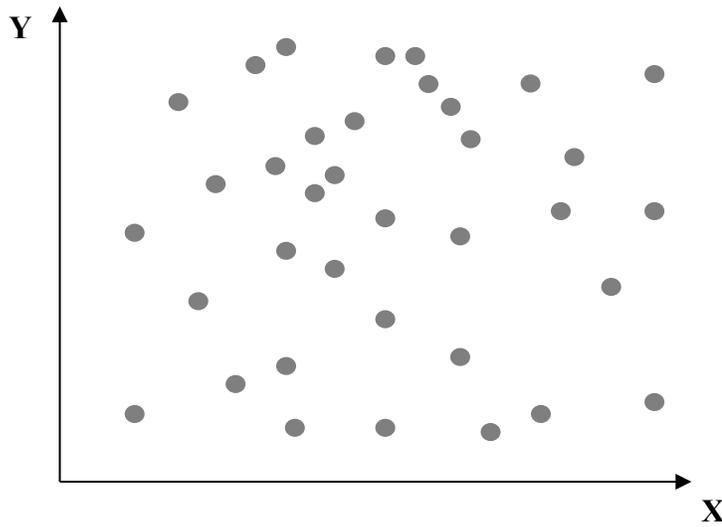
- Nguyen, C.Y., Sano, K., Tran, T.V. and Doan, T.T. (2013). Firm Relocation Patterns Incorporating Spatial Interactions, *Annals of Regional Science*, 50: 685-703.
- Openshaw, S. (1977). A Geographical Solution to Scale and Aggregation Problems in Region-Building, Partitioning, and Spatial Modelling, *Transactions of the Institute of British Geographers*, New Series 2: 459-475.
- Openshaw, S. (1984). *The Modifiable Areal Unit Problem*, Norwich: Geo Books.
- Pavlinek, P. and Zenka, J. (2011). Upgrading in the Automotive Industry: Firm-level Evidence from Central Europe, *Journal of Economic Geography*, 11(2011): 559-586.
- Ranski, H. (2011). External Economies of Localization, Urbanization and Industrial Diversity and new Firm Survival, *Papers in Regional Science*, 90(3): 473-502.
- Raspe, O. and van Oort, F. (2011). Growth of new Firms and Spatially Bounded Knowledge Externalities, *Annals of Regional Science*, 46: 495-518.
- Rathelot, R. and Sillard, P. (2008). The Importance of Local Corporate Taxes in Business Location Decisions: Evidence from French Micro Data, *The Economic Journal*, 118(March): 499-514.
- Rizov, M. and Walsh, P.P. (2011). Is There a Rural-Urban Divide? Location and Productivity of UK Manufacturing, *Regional Studies*, 45(5): 641-656.
- Rizov, M., Oskam, A. and Walsh, P.P. (2012). Is there a Limit to Agglomeration? Evidence from Productivity of Dutch Firms, *Regional Science & Urban Economics*, 42(2012): 595-606.
- Robinson, W.S. (1950). Ecological Correlations and the Behavior of Individuals, *American Sociological Review*, 15 (3): 351-357.
- Rosenthal, S.S. and Strange, W.C. (2004). "Evidence on the nature and sources of agglomeration economies," in Henderson, V. and Thisse, J.-F. (Eds). *Handbook of Regional and Urban Economics*, Volume 4: 2119-2171.
- Saito, H. and Gopinath, M. (2009). Plants' Self-selection, Agglomeration Economies and Regional Productivity in Chile, *Journal of Economic Geography*, 9(2009): 539-558.
- Shuai, X. 2013. Will specialization continue forever? A case study of interactions between industry specialization and diversity. *The Annals of Regional Science* 50:1-24.
- Strauss-Kahn, V. and Vives, X. (2009). Why and Where do Headquarters Move?, *Regional Science & Urban Economics*, 39(2009): 168-186.
- Tober, W.R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography*, 46: 234-240.
- van Soest, D.P., Gerking, S. and van Oort, F.G. (2006). Spatial Impacts of Agglomeration Externalities, *Journal of Regional Science*, 46(5): 881-899.
- Verbeek, M. and Nijman, T.E. (1992). Can Cohort Data Be Treated As Genuine Panel Data?, *Empirical Economics*, 17(1): 9-23.

Figure 1
A schematic representation of the MAUP



Legend:
● Individual firms
| | Limits of the regions

Figure 2
Location of the individual spatial units



Legend:
● Individual firms

Figure 3
Representation of construction of local indices varying the threshold distance \bar{d}

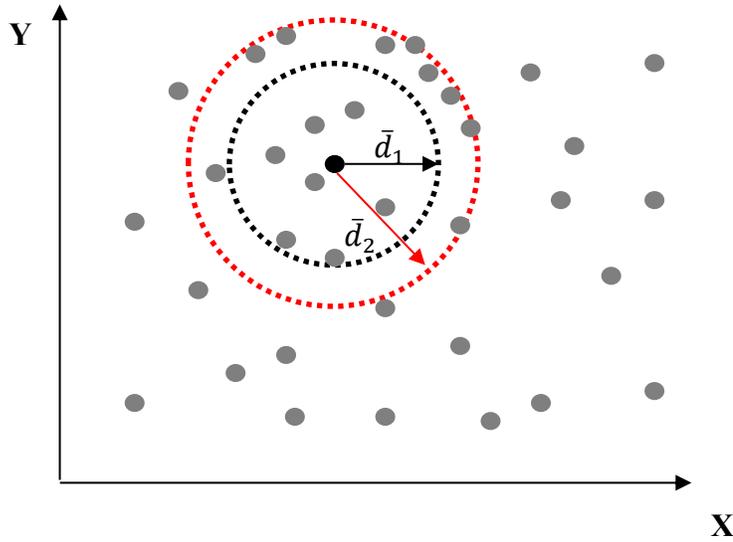


Figure 4
Map of the location of firms

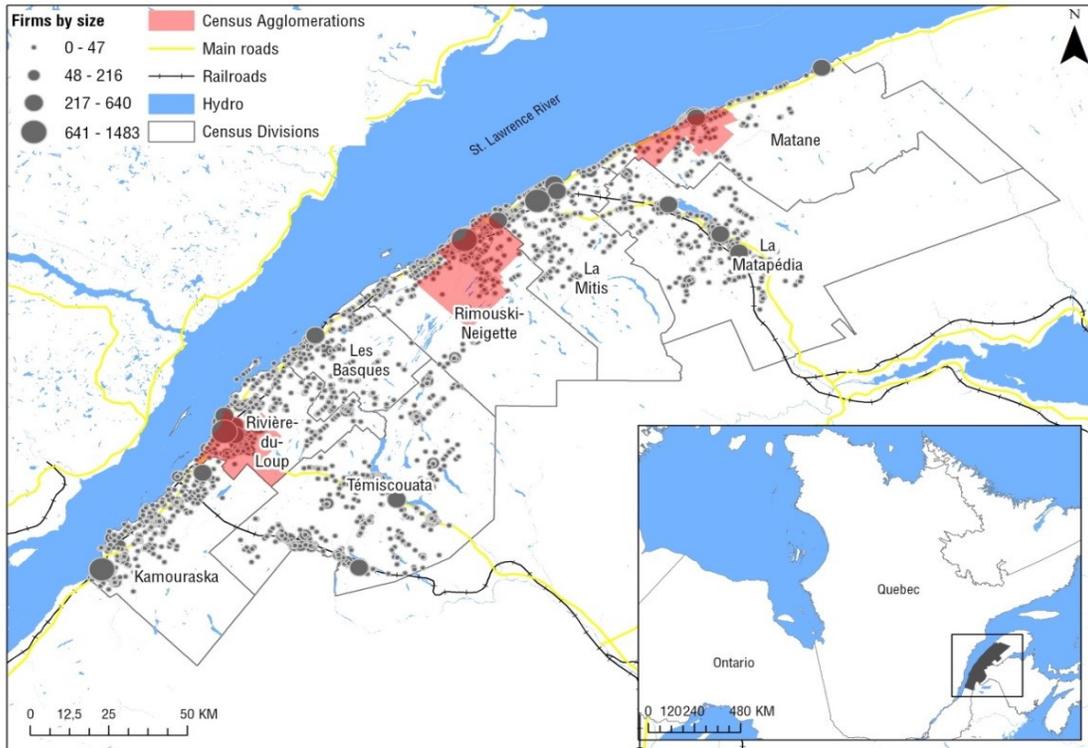


Figure 5
Distribution of the Herfindahl indices (H_k)

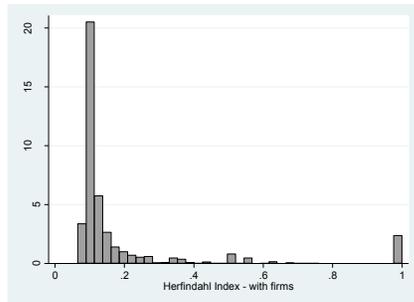


Figure 5a
Distance = 1,000 m.

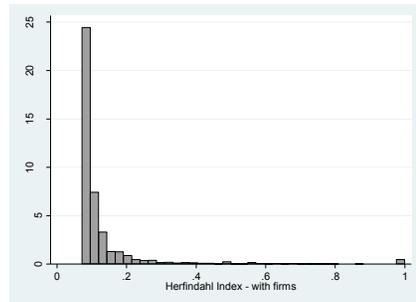


Figure 5b
Distance = 3,500 m.

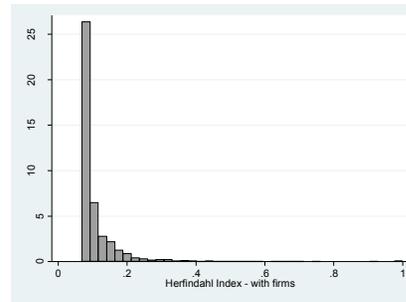


Figure 5c
Distance = 6,000 m.

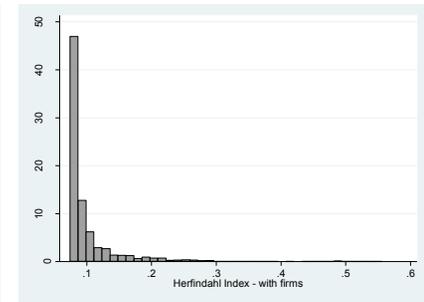


Figure 5d
Distance = 9,500 m.

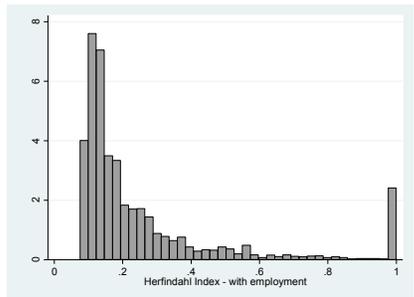


Figure 5e
Distance = 1,000 m.

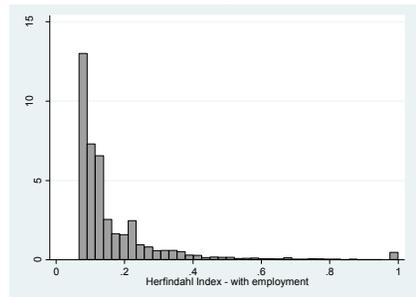


Figure 5f
Distance = 3,500 m.

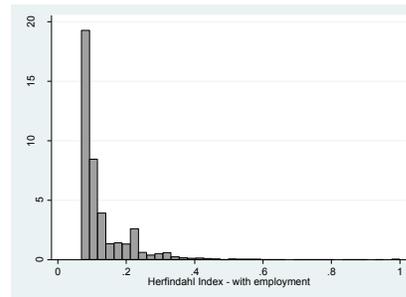


Figure 5g
Distance = 6,000 m.

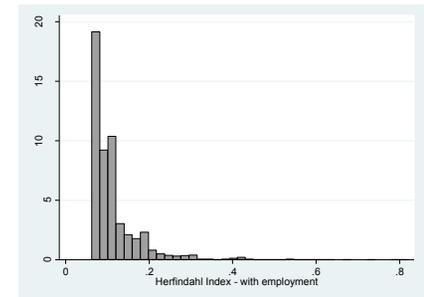


Figure 5h
Distance = 9,500 m.

Legend:

5a, 5b, 5c and 5d (number of firms)

5e, 5f, 5g and 5h (number of jobs)

Figure 6
Distribution of the Entropy indices (E_k)

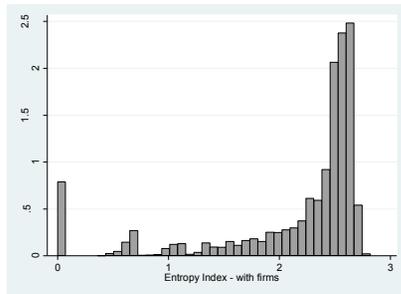


Figure 6a
Distance = 1,000 m.

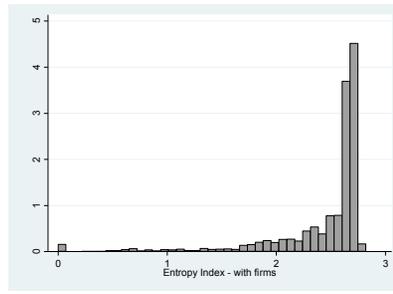


Figure 6b
Distance = 3,500 m.

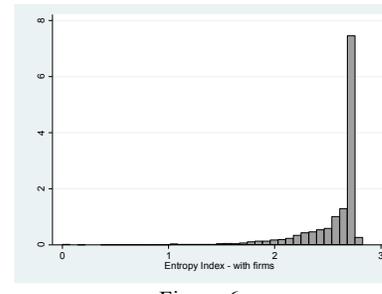


Figure 6c
Distance = 6,000 m.

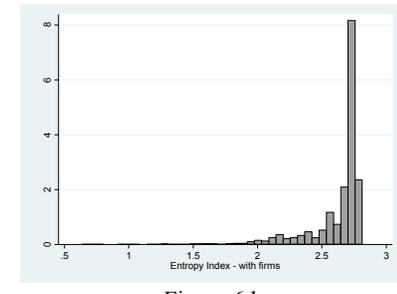


Figure 6d
Distance = 9,500 m.

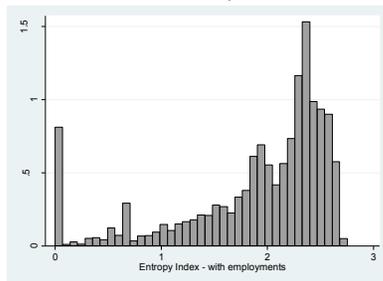


Figure 6e
Distance = 1,000 m.

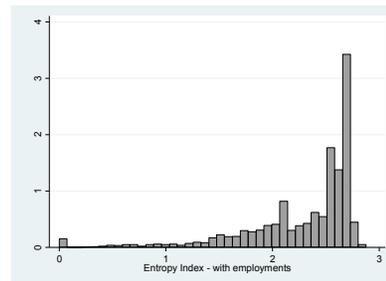


Figure 6f
Distance = 3,500 m.

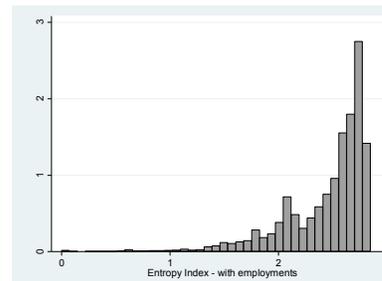


Figure 6g
Distance = 6,000 m.

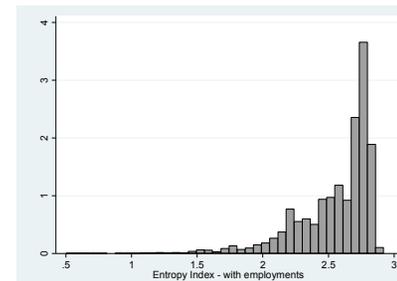


Figure 6h
Distance = 9,500 m.

Legend:

6a, 6b, 6c and 6d (number of firms)
6e, 6f, 6g and 6h (number of jobs)

Figure 7
 Distribution of the location quotient in some economic sectors (LQ_{sk})

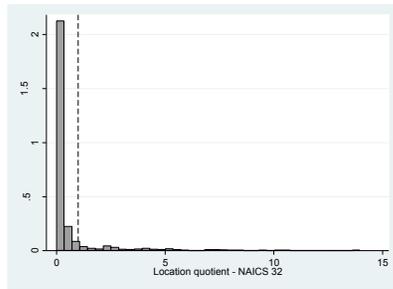


Figure 7a
 Distance = 1,000 m.

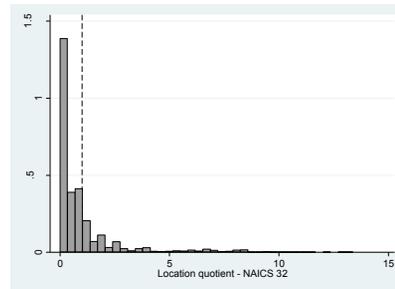


Figure 7b
 Distance = 3,500 m.

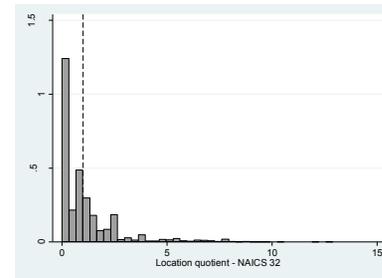


Figure 7c
 Distance = 6,000 m.

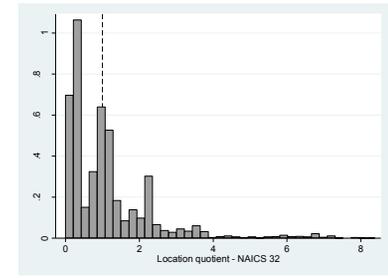


Figure 7d
 Distance = 9,500 m.

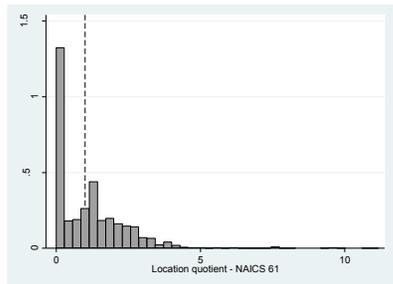


Figure 7e
 Distance = 1,000 m.

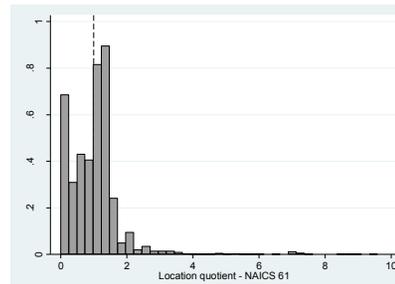


Figure 7f
 Distance = 3,500 m.

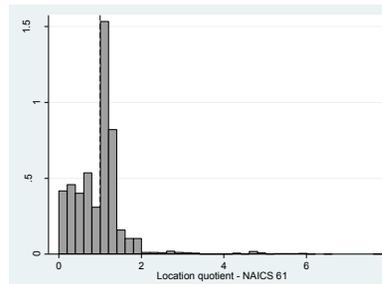


Figure 7g
 Distance = 6,000 m.

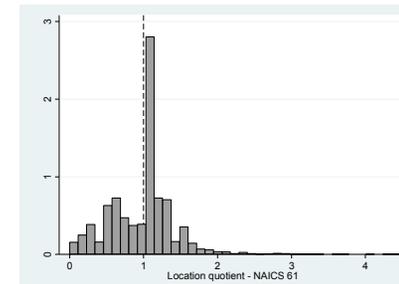


Figure 7h
 Distance = 9,500 m.

Legend:

7a, 7b, 7c and 7d: NAICS 32 (Mid-tech manufacturing)

7e, 7f, 7g and 7h: NAICS 61 (Education Services)

Table 1
List of studies that aggregate the micro data to geographical limits

Authors	Topic	number of observations (micro)	Aggregation	Place
Baldwin et al. (2012)	Economic integration	30,403,878 manufacturing plants	239 census division	Canada
Boschma et al. (2009)	Plant performance	101,093 job moves	17,098 workplaces	Sweden
Daunfeldt et al. (2013)	Start-up	51,288 firms in the wholesale industry	289 municipalities	Sweden
Eriksson and Lindgren (2009)	Firm performance	256,985 workplaces	108 functional local labour markets	Sweden
Farole and Winkler (2014)	Firm exportation	> 35,000 manufacturing and services firms	422 regions	Sweden
Figueiredo et al. (2009)	Employment size	436,092 manufacturing establishments	275 <i>concelhos</i>	Portugal
Leahy et al. (2010)	Geographical agglomeration	35,371 manufacturing establishments	1,341 Statistical local area	Australia
Leahy et al. (2010)	Geographical agglomeration	46,693 manufacturing establishments	1,336 Statistical local area	Australia
Rathelot and Sillard (2008)	Business creations	131,728 creation of firms/establishments	36,707 municipalities	France
Saito and Gopinath (2009)	Region's productivity	Manufacturing plants (food) with at least 10 employees	51 provinces	Chile

Table 2
List of different studies using micro data sets

Authors	Topic	Observations (micro data)	Place
Alama-Sabater et al. (2011)	Location choice	8,429 firms	Spain
Altomonte and Colantone (2008)	Productivity (TFP)	48,718 firms in manufacturing and construction	Romania
Andersson and Lööf (2011)	Productivity	54,278 manufacturing firms	Sweden
Aruzo-Carod and Viladecans-Marsal (2009)	Location choice	5,569 manufacturing establishments	Spain
Baldwin et al. (2010)	Urban increasing returns	11,323 manufacturing firms	Canada
Barlet et al. (2013)	Location patterns	518,036 manufacturing firms	France
Capasso et al. (2013)	Outsourcing	205 manufacturing firms	Italy
De Vaan et al. (2013)	Survival patterns	4,607 firms in video game industry	Sweden
Demirel and Mazzucato	Firm growth	256 pharmaceutical firms	USA
Figueiredo et al. (201)	Agglomeration economies	209,149 establishments	Portugal
Fontagné et al. (2013)	Cluster policy	111,960 exporting firms	France
Giunta et al. (2012)	Firm growth	7,878 manufacturing firms	Italy
Graham and Kim (2008)	Productivity (TFP)	74,346 manufacturing and services firms	UK
Heebels & Boschma (2011)	Survival firms	1,434 publishing firms	Netherlands
Holl (2012)	Productivity	2,470 manufacturing firms	Spain
Holl et al. (2010)	Subcontracting	162 companies in electronic industry	Spain
Ibrahim et al. (2009)	Clusters	165 inventors in telecommunication	USA
Koenig	Export decision	5,776,140 manufacturing firms	France

Table 2 (end)
List of different studies using micro data sets

Authors	Topic	Observations (micro data)	Place
Kronenberg (2013)	Relocation decision	179,913 individual firms	Netherlands
Lee (2008)	Development policy	268,367 manufacturing plants	USA
Lopez and Sudekum (2009)	Productivity (TFP)	40,454 manufacturing establishments	Chile
Mota and Brandao (2011)	Location choice	61,177 plants births	Portugal
Nakamura (2012)	Productivity	44,020 firms	Japan
Nguyen et al. (2013)	Relocation process	3,810 firms	Tokyo (Japan)
Pavlinek and Zenka (2011)	Industrial upgrading	490 firms	Czech
Ranski (2011)	Survival of firms	31,069 firms	USA
Raspe and van Oort (2011)	Firm growth	28,701 manufacturing and business establishments	Netherlands
Rizov and Walsh (2011)	Productivity	6,722 manufacturing firms and 23,841 establishments	UK
Rizov et al. (2012)	Productivity	4,220 small and medium firms (13,897 establishments)	Netherlands
Strauss-Kahn and Vives (2009)	Location decision	26,195 headquarters	USA

Table 3
Distribution of the firms according to the region (Figure 4) and the economic sector

	Rimouski- Neigette	Rivière- du- Loup	Matane	Les Basques	La Mitis	La Matapedia	Kamouraska	Témiscouata	Rimouski (RMR)	Rivière- du- Loup (RMR)	Matane (RMR)	Total
Agri., forest., & fishing (11)	61	113	54	170	153	152	196	190	102	32	46	1,269
Mining, oil & gas extraction (21)	4	2	0	1	0	2	4	2	2	8	3	28
Public services (22)	0	0	0	0	1	0	1	1	2	3	0	8
Construction (23)	19	35	13	39	53	60	69	54	139	93	52	626
Low-tech manufacturing (31)	8	6	2	7	13	4	18	6	10	12	13	99
Mid-tech manufacturing (32)	6	11	5	15	20	23	16	44	20	20	10	190
Mid/high-tech manufacturing (33)	7	10	2	10	12	12	20	23	42	35	13	186
Wholesale trade (41)	8	5	1	15	16	16	19	16	100	36	16	248
Retail trade - General (44)	22	45	24	62	107	100	134	123	309	198	122	1,246
Retail trade - Dpt. stores (45)	7	13	5	17	32	21	49	33	111	62	47	397
Transportation (48)	10	16	7	24	40	40	37	45	50	60	29	358
Warehousing (49)	2	6	4	6	10	7	12	11	7	7	4	76
Inf. & cult. Industries (51)	1	4	1	5	6	8	22	10	38	22	16	133
Finance & Insurance (52)	7	9	6	13	23	31	32	39	79	44	33	316
Real Est. & Leasing (53)	3	1	0	4	13	10	15	8	64	32	22	172
Prof. Sci. & Tech. Serv. (54)	9	9	3	19	29	30	64	41	167	107	30	508
Mgmt. of Comp. (55)	0	0	1	1	0	0	1	0	5	4	0	12
Adm. Supp. & Waste Services (56)	2	5	0	8	12	12	18	7	67	32	20	183
Education Services (61)	7	9	9	16	23	24	30	35	45	29	10	237
Health & Social Ass. (62)	12	24	4	39	66	61	77	88	185	104	70	730
Arts, Entertainment & Rec. (71)	9	7	5	19	12	16	32	29	50	26	19	224
Accommodation & Food Serv. (72)	25	29	25	46	67	64	79	105	139	119	53	751
Other Services (81)	39	70	22	81	129	126	182	185	347	181	121	1,483
Public Adm. (91)	10	16	14	21	36	36	40	45	83	30	28	359
Total	278	445	207	638	873	855	1,167	1,140	2,163	1,296	777	9,839

Table 4

Distribution of the firms according to the region (Figure 4) and the size of the firms

	Rimouski- Neigette	Rivière- du- Loup	Matane	Les Basques	La Mitis	La Matapedia	Kamouraska	Témiscouata	Rimouski (RMR)	Rivière- du- Loup (RMR)	Matane (RMR)	Total
0 to 5 employees	222	361	173	510	657	671	884	885	1,395	896	548	7,202
6 to 10 employees	28	44	17	61	104	71	127	117	356	169	102	1,196
11 to 15 employees	14	20	6	29	36	37	57	48	142	64	40	493
16 to 20 employees	3	2	4	10	18	19	26	28	77	39	33	259
21 to 25 employees	3	4	1	10	13	20	13	12	36	23	9	144
26 to 50 employees	6	9	4	11	22	20	34	23	98	58	21	306
51 to 75 employees	1	3	0	4	8	5	9	12	23	12	8	85
76 employees or more	1	2	2	3	15	12	17	15	36	35	16	154
Total	278	445	207	638	873	855	1,167	1,140	2,163	1,296	777	9,839

Table 5

Descriptive statistics of the local indices according to the number of employees/establishment- Herfindahl (H_k) and entropy (E_k)

	Herfindhal Index		Entropy Index	
	Mean	Standard deviation	Mean	Standard deviation
500 m.	0.2272	0.2372	1.6613	0.7097
1,000 m.	0.1924	0.2209	1.9031	0.7124
1,500 m.	0.1759	0.2046	2.0150	0.7020
2,000 m.	0.1611	0.1838	2.1025	0.6655
2,500 m.	0.1497	0.1644	2.1683	0.6278
3,000 m.	0.1394	0.1437	2.2308	0.5876
3,500 m.	0.1311	0.1227	2.2839	0.5419
4,000 m.	0.1245	0.1057	2.3286	0.4993
4,500 m.	0.1202	0.0937	2.3620	0.4691
5,000 m.	0.1170	0.0852	2.3932	0.4426
5,500 m.	0.1136	0.0731	2.4200	0.4135
6,000 m.	0.1117	0.0667	2.4404	0.3918
6,500 m.	0.1098	0.0593	2.4610	0.3700
7,000 m.	0.1083	0.0548	2.4805	0.3524
7,500 m.	0.1075	0.0530	2.4934	0.3467
8,000 m.	0.1066	0.0505	2.5059	0.3398
8,500 m.	0.1054	0.0489	2.5218	0.3254
9,000 m.	0.1042	0.0475	2.5362	0.3098
9,500 m.	0.1033	0.0463	2.5487	0.2993
10,000 m.	0.1024	0.0442	2.5603	0.2884