ÁREA TEMÁTICA: 1. ECONOMIA

RELATED INDUSTRIES, ECONOMIC COMPLEXITY, AND REGIONAL DIVERSIFICATION: AN APPLICATION FOR BRAZILIAN MICROREGIONS

Elton Eduardo Freitas[1]
PhD at Cedeplar-UFMG
Av. Antônio Carlos 6627 - Prédio FACE - Campus Pampulha
Belo Horizonte – MG - 31270-901 - Brazil
+553134097070

Gustavo Britto
Professor at Universidade Federal de Minas Gerais, Department of Economics
Researcher at Cedeplar-UFMG
Av. Antônio Carlos 6627 - Prédio FACE - Campus Pampulha
Belo Horizonte – MG - 31270-901 - Brazil
+553134097070

Pedro Amaral
Professor at Universidade Federal de Minas Gerais, Department of Economics
Researcher at Cedeplar-UFMG
Av. Antônio Carlos 6627 - Prédio FACE - Campus Pampulha
Belo Horizonte – MG - 31270-901 - Brazil
+553134097070

ABSTRACT

This paper contributes to the conceptual and empirical literature on regional diversification as a process of related industrial diversification. We developed a new measure of relatedness measure between economic activities. The empirical exercise shows that productive specialization of regions is a strongly path dependent process, where new economic activity is conditioned by the already existing productive structure. The analyzes carried out suggest that it is difficult to attract new industries to a region if they are technologically unrelated from existing local activities. This difficulty becomes even greater in the case of complex industries.

Keywords: Economic complexity; Industry relatedness; Regional diversification; Technological cohesion

JEL classification: O25, O38, R11

Acknowledgements: The authors are grateful for financial support received from FAPEMIG, CAPES and CNPq.

[1] To whom correspondence should be addressed.

1. INTRODUCTION

What determines regional productive diversification and how regions develop new paths of growth is a longstanding research question that is a large body of the economic literature. More recently, studies have breathed new life into the theoretical and applied literature showing that new local industries evolve from regional industrial structures that provide skills and assets (Frenken *et al.*, 2007, Boschma and Iammarino, 2009, Neffke *et al.*, 2011, Boschma *et al.*, 2013, Rigby, 2013; Essletzbichlera, 2015, Balland *et al.*, 2018).

Long before, during the 1990s, the bulk of the research on this field focused on the degree of relatedness between sectors. Clusters of industries were identified on the basis of their technological complementarities and various methods for measuring the technological relationship between industries were developed and applied (see, *e.g.*, Farjoun, 1994; Teece *et al.*, 1994). In parallel, the economic relevance of urban diversity and its effects on agglomeration economies was empirically evaluated, suggesting that local competition and urban diversity could promote the growth of industrial employment, in what was branded as the power of Jacobs externalities (Glaeser *et al.,* 1992; Ellison and Glaeser, 1997).

This branch of the literature opened way for research on different types of externalities and their role on sustaining traditional industries and attracting new industries. As empirical literature has suggested, new industries, particularly high-tech industries, developed in diversified cities where Jacobs externalities were available, while mature industries benefited most from location externalities that were generated in more specialized cities (Henderson *et al.,* 1995; Brezis and Krugman, 1997). Jacobs (1969) argued that the largest and most relevant source of externalities is the diversity of economic activities that is developed in cities. In this case, the multiplicity of goods and services, technologies and own knowledge, which has a diversified urban center, enhances what is referred to as the cross fertilization of ideas, that is, innovations originated from the fertilization of ideas among the various sectors of activities that are housed in the same city and driven by the generation of new types of jobs. These newly developed jobs increase the capacity for generating new goods and services.

In the 2000s, the idea of industry relationship was combined with the empirical observation made by economic geographers that knowledge spillovers were often geographically limited. There is evidence that the variety of industries or technologies present in one region can positively affect knowledge and learning because local firms in different (but related) activities can profit more from mutual spillovers than from local firms in unrelated industries (Almeida and Kogut, 1999; Boschma and Frenken, 2009, 2011; Gilsing *et al.*, 2007; Menzel, 2008, Balland *et al.*, 2018).

Thus, the greater the variety among related sectors in a region, the more opportunities for learning exist for local industries, which makes intersectoral overflows of knowledge more likely and causes greater economic performance of these regions. Some recent studies have found empirical support for the importance of the related variety for regional growth in the Netherlands (Frenken *et al.*, 2007), Italy (Boschma and Iammarino, 2009), Sweden (Neffke *et al.*, 2011; Boschma *et al.*, 2013), United States (Rigby, 2013; Essletzbichlera, 2015) and the Europe (Balland *et al.*, 2018).

In addition to the tenet that the relationship between industries in a region can foster regional growth, there is evidence that territories are more likely to expand and diversify towards sectors that are closely related to their existing activities (Hausmann and Klinger, 2007; Hidalgo *et al.*, 2007; Hausmann and Hidalgo, 2010). These recent studies argue that a country's current productive structure affects its future state because the existing pool of capabilities in a country determines which new industries will be developmentally viable during the near future.

Countries that accumulate larger sets of capabilities tend to produce more specialized products that are difficult to copy or imitate by others (Hidalgo and Hausmann, 2009). The complexity of an economy is incorporated into the wide range of knowledge or capabilities that are combined to make products; less ubiquitous products are more likely to require a greater variety of resources. These specialized (complex) goods tend to be produced by relatively few economies and form the basis for long-term competitive advantage.

Hence, it would befit regional policy to build comparative advantages in complex technologies (Hidalgo and Hausmann, 2009; Hausmann *et al.*, 2011). Once a region succeeds, it can grow further in these technologies based on accumulated technological advantages. However, complex technologies are relatively scarce, which makes it difficult for regional economies to develop skills in these fields. These two tendencies give rise to a dilemma of diversification that is caused by the trap of low complexity (Hausmann and Hidalgo, 2011). On the one hand, it is not possible to make new products because they do not have the necessary capabilities. On the other hand, it is not possible to accumulate capabilities because the products that need them are not produced; that is, complex technologies remain out of reach for most because they lack the diversity of capabilities from which complex technologies are derived.

The general solution to this dilemma is for regional economies to develop their existing knowledge centers and expand their technological repertoires along related trajectories that lead to more complex technologies. Thus, the emergence of new technologies and new sectors within the regions is not random but reflects the existing collective capability of agents that produce in regions that have different technological and industrial profiles.

Hence, there is an untapped potential to shed light on the connections between industrial relatedness (diversification) and agglomeration economies by applying and developing the growing theoretical and applied insights of economic complexity to firm-level, regional data. This paper contributes to this debate in three ways. First, by connecting the literature on agglomeration economies (Glaeser *et al.*, 1992; Henderson *et al.*, 1995) with that of regional diversification (Hidalgo *et al.*, 2007; Neffke *et al.*, 2011, Balland *et al.*, 2018), and with that of economic complexity (Hidalgo and Hausmann, 2009; Hausmann *et al.*, 2011). Secondly a new measure of relatedness measure between economic activities is developed. Finally, it provides empirical evidence for the process of diversification on a regional, rather than on a national scale. To do so, industries that enter and exit a given region are compared to those that were already present. The underlying assumption is that relationships and economic complexity are key components of related diversification. We consider this to be a process through which regions' economic structures are improved based on their existing capabilities (Boschma, 2014).

Based on the theoretical an applied literature, we formulated four hypotheses regarding the related diversification process that will be tested for the Brazilian microregions:

> Hypothesis 1: regions are more likely to develop new specializations in technological activities related to their productive structures.
>
> Hypothesis 2: regions are less likely to develop new specializations in technological activities unrelated to their productive structures.
>
> Hypothesis 3: regions are less likely to develop new specializations in complex technological activities.
>
> Hypothesis 4: regions are more likely to develop new specializations in complex technological activities when their productive structures are also complex.

To test this hypotheses, we have developed a relatedness measure between economic activities and use it to show that the process of regional diversification is conditioned by existing regional productive structures.

The article is structured as follows. The following section presents the distinct perspectives considered in the paper. The third section presents the main results of the econometric analysis. The last section presents the main conclusions.

2. RELATEDNESS, DENSITY AND ECONOMIC COMPLEXITY

In this section the relevant literature in economic density, relatedness, and complexity are brought together by simultaneously assessing the existing measures and indexes and developing new ones.

To do so, the study uses identified microdata that are part of the collection of administrative records of the Annual Report on Social Information (RAIS) of the Ministry of Economy of Brazil that were taken between 2006-2016. RAIS provides yearly data for all Brazil municipalities comprising all formally registered firms and their employees. The dataset contains information on several characteristics of the firm, such as sector of activity and firm size, as well as the employees, such as individual wage, occupation, age, sex, among other.

This work uses aggregate sector data according to the National Classification of Economic Activities (CNAE 2.0), composed of 1,329 subclasses (7 digits level for industries) aggregated into 21 Sections (one digit level for industries). We focus on the following activity sectors: transformation industry, extractive industry, agricultural and livestock sectors and productive and distributive services. This means that the level of disaggregation used is the highest possible in the CNAE classification. Table A.1 (see Appendix) presents the selected sectors and their respective division codes.

2.1. RELATEDNESS MEASURED AS CO-OCCURRENCE

In empirical works, the relatedness between industries is captured in a certain way by the hierarchical structure of the industrial classification adopted. These classifications are already designed to place the closest industries within the same classification system. The smaller the class that two industries share in the hierarchy of industrial classification, the more similar they are. According to this logic, industries of the same 5-digit class are more related than, for example, industries that share only the same 2-digit class. However, this measure has been questioned because it is quite rigid and devoid of theory (Teece *et al.*, 1994; Bryce and Winter, 2009; Neffke *et al.*, 2011).

A number of alternative approaches have emerged because of the limitations of measures that are based on the hierarchy of industrial classifications. For a long time, the most influential approach was the Scherer (1982) method, which constructs a matrix in which related industries (based on technology) flow. In the 1990s, Engelsman and Van Raan (1991, 1994) used the fact that some patents are classified into various classes of technology as evidence for the technological relationship between these classes. Patents are classified by at least one classification code (primary or primary) of the International Patent Classification, but generally, more classification codes (secondary or supplementary) are assigned to patents. The assumption is that the frequency with which two classification codes are jointly assigned the same patent can be interpreted as a sign of the strength of the knowledge relationship between the technological fields that the codes represent, i.e. as a measure of proximity between the base knowledge of the two fields.

Other measures of industrial relatedness were developed and based on the similarities of the input-output chains of the sectors (Fan and Lang, 2000), or on the similarities in the mix of occupations employed by different industries (Farjoun, 1994).

More recently, several scholars have used co-occurrence analyses to assess the relationship between industries (Teece *et al.,* 1994; Bryce and Winter, 2009; Hidalgo *et al.*, 2007). Co-occurrence analysis measures the degree of consistency by assessing whether two industries are often found together from the same economic unit of analysis. For example, in Hidalgo *et al.* (2007), the number of times that two sectors had a comparative advantage were counted, which revealed (co-occurrence) in the same country (the economic unit). Likewise, Teece *et al.* (1994) and Bryce and Winter (2009) count the number of times a company (the economic unit) has industrial plants in two different sectors (co-occurrence). However, other factors may influence the number of co-occurrences in addition to the degree of consistency. For example, a very large sector is more likely to present a larger number of firms and therefore, they co-occur more frequently with other industries.

Neffke and Henning (2008) developed a measure that is based on co-occurrence and used this measure to estimate the degree of relatedness, called revealed relatedness. Industrial relatedness is derived from the co-occurrence of products belonging to the portfolio of different industrial plants. The central hypothesis is of two products are produced in the same plant, there may be relatedness between the industries that these products are a part of.

Hidalgo *et al.* (2007) and Hausmann and Klinger (2007) developed what they call Product Space. A system in which similar products are connected based on the likelihood of being co-exported. The calculation of similarities between products is based on the concept of Revealed Comparative

Advantage (RCA)[2], i.e. if RCA ≥ 1, we have that the country is an effective exporter of a given well p, but for RCA <1, the country is not competitive. This is described in Equation 1.

$$RCA_{p,c} = \frac{X_{p,c}/X_p}{X_c/X} \tag{1}$$

where: $X_{p,c}$ is the exported value of a product *p* by a country *c*; X_p is the worldwide exportation of the product *p*; X_c is total export of a country *c*; X is the total value of world exports.

After calculating this indicator, Hausmann and Klinger (2007) and Hidalgo *et al.* (2007) developed a methodology that uses conditional probabilities to establish connections between products. Probabilities of exporting a particular product because another product is exported are calculated for each product. These probabilities, called 'proximity' by the authors, are then used to determine the strength of the bonds between the different products. Equation 2 presents the measure of proximity between two products *p* and *p'*. This was originally developed by Hidalgo *et al.* (2007).

$$\varphi_{p,p'} = min\{P(RCA_p|RCA_{p'}), P(RCA_{p'}|RCA_p)\} \tag{2}$$

where for every country *c*:

$$RCA_{p,c} = \begin{cases} 1, \text{ if } \mathrm{RCA}_{\mathrm{p,c}} \geq 1 \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

In order to estimate the relationship between industries, we take as a starting point the co-occurrence approach and adapt the proximity indicator (Equation 2), constructed by Hidalgo *et al.* (2007) based on employment data and economic activities of RAIS for the period between 2006 and 2016. Relationships between industries will be captured through three dimensions. The first dimension, called co-occupation, will focus on the similarities in the mix of occupations employed by different industries to capture the relationship between them. The second dimension, called co-localization, will focus on the number of times that two sectors appear in the same micro-region. Finally, the third dimension, which we will call co-company, will count the number of times a company has industrial plants in two different sectors.

First, to construct the method that will compare the similarities between the occupations employed by different industries, we adapted the calculation of the RCA (Equation 1) to capture effective occupations in each industry as follows:

$$EO_{i,o} = \frac{emp_{i,o}/emp_i}{emp_o/emp} \tag{4}$$

---

[2] See Balassa (1965).

where: $emp_{i,o}$ is the employment of the occupation o in the sector i; emp_i is the total employment of the sector i in the country; emp_o is the total employment of the occupation o in the country; emp is total employment in the country.

Thus, when EO is greater than unity indicates that the share of an occupation in the employment of an industrial sector is greater than the share of that occupation in national employment. When OE is greater than or equal to 1, we say that the sector i actually employs occupation o, and when OE is less than 1, that sector is not an effective employer of that occupation.

We substitute EO in Equation 2 to calculate the probability of an industry employing a particular occupation because this occupation is employed in another industry. Equation 5 presents the measure of co-occupation between two industries i and j.

$$\theta_{i,j} = min\{P(OE_{i,o} = 1|OE_{j,o} = 1), P(OE_{j,o} = 1|OE_{i,o} = 1)\},\ \forall\ i \neq j \tag{5}$$

in which for all industry i:

$$EO_{i,o} = \begin{cases} 1, \text{ if } EO_{i,o} \geq 1 \\ 0, \text{ otherwise} \end{cases} \tag{6}$$

For the co-localization dimension, we will use the Locational Quotient (LQ), a measure considered in this work as a proxy for industrial specialization. The indicator of local sectorial specialization is a measure of industrial concentration and measures the degree of specialization of each sector in each of the analyzed regions. The LQ is described as:

$$LQ_{m,i} = \frac{emp_{m,i}/emp_m}{emp_i/emp} \tag{7}$$

where: $emp_{m,i}$ is the employment of industry i in the microregion m; emp_m is the total employment of the micro-region m; emp_i is the total employment of the industry i in the country; employment in the country.

This reflects the fraction of employees of a given industry, in a given location, relative to the total employee fraction of the industry over the total employment level. If the calculated LQ indicator is greater than unity, then the micro-region m has a high participation of sector i compared to the relative proportion of the other microregions.

We substitute LQ in Equation 2 to calculate the probability of an industry being co-located with another industry. Equation 8 presents the measure of co-location between two industries i and j.

$$\phi_{i,j} = min\{P(LQ_{m,i} = 1|LQ_{m,j} = 1), P(LQ_{m,j} = 1|LQ_{m,i} = 1)\},\ \forall\ i \neq j \tag{8}$$

in which for every microregion m:

$$LQ_{m,i} = \begin{cases} 1, \text{ if } LQ_{m,i} \geq 1 \\ 0, \text{ otherwise} \end{cases} \quad (9)$$

For the co-company dimension, we assume that industries that are more related will be more often be found in the same corporation (Penrose, 1959; Teece *et al.*, 1994; Bryce and Winter, 2009). Thus, if firms that participate in industry *i*, in general, also participate in industry *j*, we can conclude that these activities are related. Consequently, industries that rarely or never appear combined are unrelated. To do so, we define an adjacency matrix $M_{f,i}$ to summarize the company that participate in one or more industries, such as:

$$M_{f,i} = \begin{cases} 1, \text{ if a company } f \text{ participate in an industry } i \\ 0, \text{ otherwise} \end{cases} \quad (10)$$

The total number of firms participating in industry *i* is given by $m_i = \sum_f M_{f,i}$, that is, the sum of $M_{f,i}$ over all firms. The number of industries in which a firm participates is given by $n_f = \sum_i M_{f,i}$. Now consider the number of firms that participate in both *i* and *j* industries, such as:

$$K_{i,j} = \sum_f M_{f,i} M_{f,j} \quad (11)$$

This count of joint occurrences of industries, $K_{i,j}$, can be used to construct a relationship measure. We redefined Equation 2 to calculate the probability of a co-occurrence of two industries *i* and *j*. Equation 12 presents the co-company measure between two industries *i* and *j*.

$$\lambda_{i,j} = min\{P(I|J), P(J|I)\} \quad (12)$$

where $P(I|J) = K_{i,j}/m_j$ and $P(J|I) = K_{i,j}/m_i$ indicate the probability that a company is participating in industry *i* if it participates in industry *j*.

These three indicators provide the necessary parameters for the elaboration of a single industrial relationship indicator, which we will call the Relatedness Measure (*RM*). For its calculation, we propose to make a linear combination made from the average of the three indicators for the period between 2006 and 2016 (Equation 13).

$$RM_{ij} = \alpha_1 \theta_{i,j} + \alpha_2 \phi_{i,j} + \alpha_3 \lambda_{i,j} \quad (13)$$

To obtain the weights (α) of each of the indexes defined in Equation (13), we used a multivariate principal component analysis (PCA). Thus, Equation (14) presents the specific weights for each indicator that take into account their participation in explaining the potential for the relationship between industries.

$$RM_{ij} = 0{,}33 \cdot \theta_{i,j} + 0{,}30 \cdot \phi_{i,j} + 0{,}37 \cdot \lambda_{i,j} \quad (14)$$

2.2. MEASUREMENT OF REGIONAL RELATEDNESS

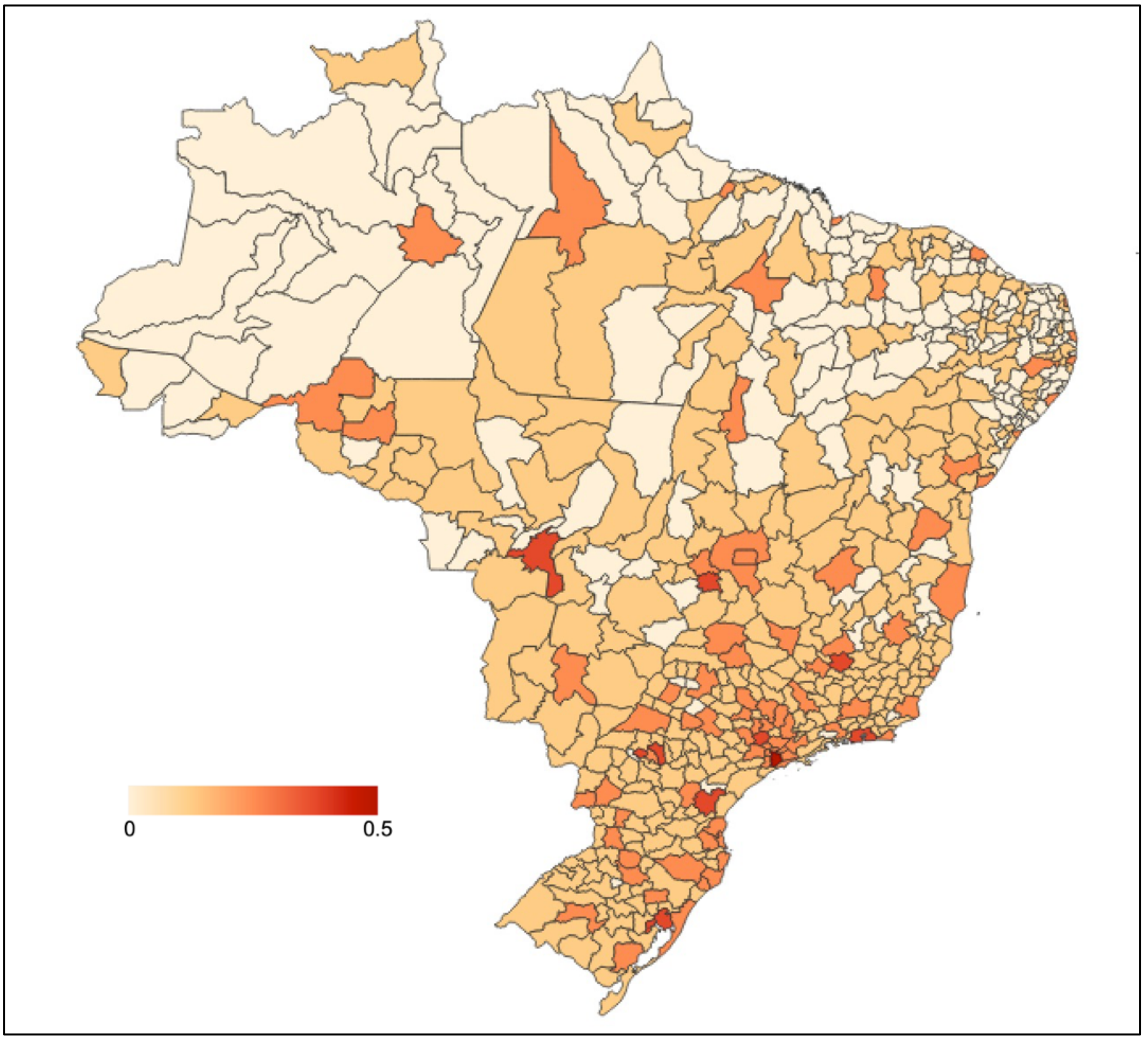
Hausmann *et al.* (2011) constructed an index to calculate the distance of an industry to the productive structure of a region. This measure, called Density, measures the distance between a given sector in relation to the productive structure of a region, also meaning the difficulty of this region in specializing in a given industry. The idea here is that each sector requires a set of productive knowledge, which may or may not be shared with other sectors (given the proximity). Sectors closer to others, for which the region is already specialized, will be more easily developed - precisely because some of the necessary knowledge is already present in the region. Formally,

$$d_{m,i,t} = \frac{\sum_{i'} M_{mit} RM_{ii'}}{\sum_{i'} CM_{ii'}} \quad (15)$$

where: $RM_{ss'}$ is the proximity of sector i to a sector i' which is defined by the Relatedness Measure presented in Equation 14; M_{mst} is a matrix that indicates whether the region is specialized or not in the employment of each sector i in a given year t. The M_{mit} is a binary integer matrix, that is, containing values 0 or 1, which assumes the value 1 when a region is specialized in a sector i, that is, when it has a high participation of employment in sector i compared to the proportion of the other regions and is otherwise 0. We will use the Locational Quotient (*LQ*), measure was described in Equation (7), as a proxy for industrial specialization.

Figure 1 shows the mean densities between all industries and each microregion for the 2006-2016 period for all Brazilian microregions. The higher the average density of a micro-region, the closer to the existing set of industries the industries that are lacking in the region are. In other words, it reflects an overall average score of a region's potential to develop new sectors. Figure 1 also shows that there are huge differences in the potential for diversification among Brazilian regions. In general, the Southeast and the area South of the country have great potential for developing new sectors. This is in contrast to many microregions in the Northeast and North, where the opportunities for diversification are much smaller.

FIGURE 1: Average density of the Brazilian microregions in the period 2006-2016.



Source: Authors' elaboration based on data from RAIS.

2.3. MEASUREMENT OF ECONOMIC COMPLEXITY

We developed two complexity indexes based on the "method of reflections" developed by Hidalgo and Hausmann (2009), Industry Complexity Index (ICI) and Economic Complexity Index (ECI). In their pioneering work, Hidalgo and Hausmann show that the economic complexity of a country's output is reflected by the particular composition of its export basket, taking into account the relative composition of the export baskets of all other countries. The main idea in their analytical framework is that more complex economies produce more exclusive goods, i.e. non-ubiquitous commodities that are sourced in relatively few countries in total. Countries with complex economic structures experience a privileged source of comparative advantage, a form of spatial-technological-monopoly from which they extract rents. Countries that produce goods that are widely imitated by others, commodities that are ubiquitous, tend to have low scores in terms of economic complexity.

Following this approach, we analyze the particular architecture of the regions-industry network and we show that a micro-region has a complex industry composition if it produces the that relatively few other regions are able to imitate. We consider only the microregions that are significant producers in certain types of industry; we consider the economic activities for which the regions have *LQ* greater than 1 in a certain period. As a result, the adjacency matrix elements that we examine M_{mi} reflect whether the microregion *m* is specialized or not in economic activity *i*.

From the concepts described above, we can evaluate the complexity of a region using the matrix M_{mi} composed of 1 (one) if the microregion *m* has *LQ*> 1 for a sector *i*, and 0 (zero) otherwise. Then, the diversity and ubiquity are calculated from equations (16) and (17), respectively.

$$K_{m,0} = \sum_i M_{mi} \tag{16}$$

$$K_{i,0} = \sum_m M_{mi} \tag{17}$$

We follow Hidalgo and Hausmann (2009) and sequentially combine equations (16) and (17) computing simultaneously the following 2 equations over a series of n iterations:

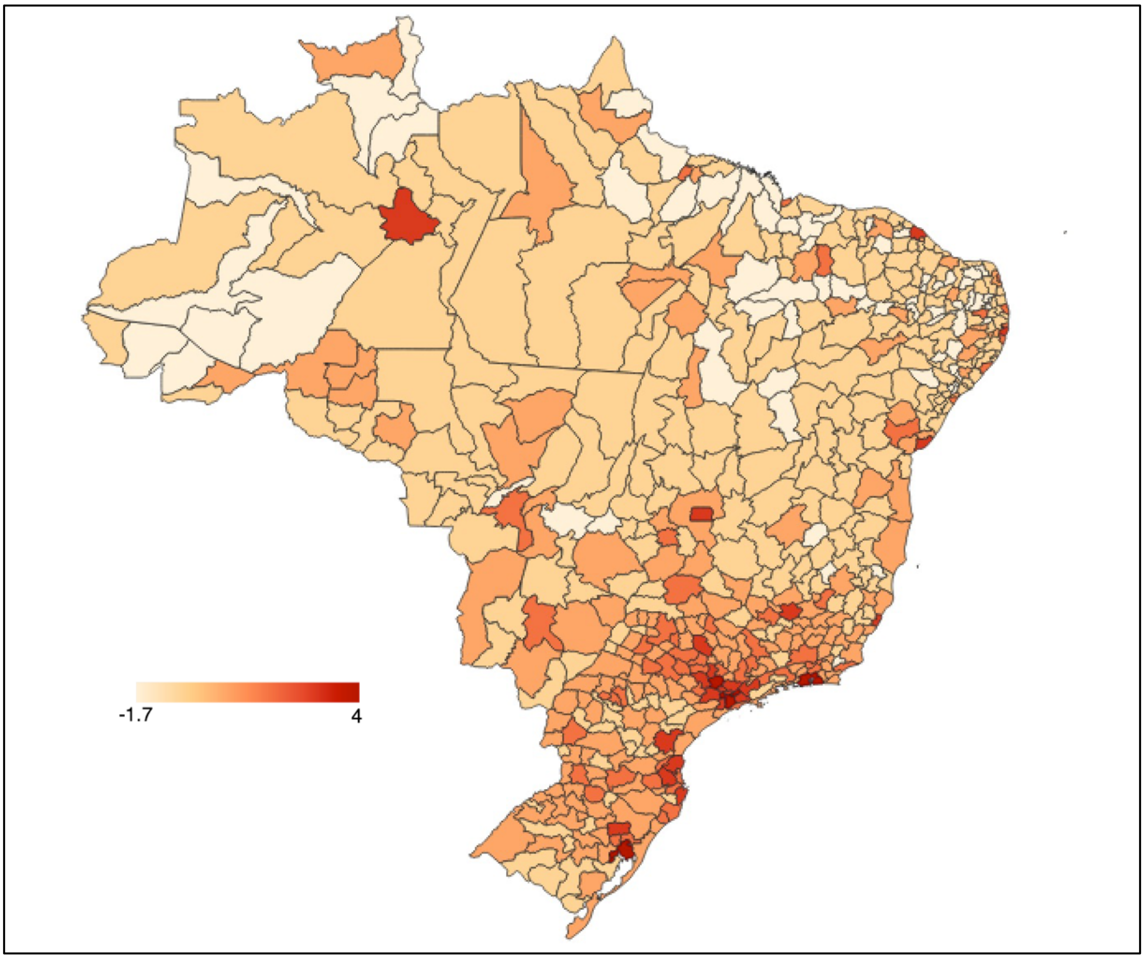
$$K_{m,N} = \frac{1}{K_{m,0}} \sum_i M_{mi} . K_{i,N-1} \tag{18}$$

$$K_{i,N} = \frac{1}{K_{i,0}} \sum_m M_{mi} . K_{m,N-1} \tag{19}$$

To provide some further interpretation of this method, in a second iteration, for N = 1, $K_{m,1}$ in equation (18) represents the average ubiquity of the industries in which region *m* has *LQ*> 1. In similar fashion, $K_{i,1}$ in equation (19) measures the average diversity of regions that have *LQ*> 1 in industry *i*. In the next iteration, N = 2, $K_{m,2}$ in for captures the average diversity of regions that have export baskets similar to region *m*, and $K_{i,2}$ reveals the average ubiquity of the industries in cities that have *LQ*> 1 in industry i. Each additional step in yields a finer-grained estimate of the industry complexity of a region using information on the complexity of the industries in which the region exhibits *LQ*> 1. Each additional step in yields a finer-grained estimate of the region complexity using information on the complexity of the industries in which the region exhibits *LQ*> 1. Each additional step in provides a finer-grained estimate of the industry complexity using information on the complexity of regions that have *LQ*> 1 in that industry. While higher order iterations in this technique become progressively more difficult to define, the method of reflections provides more and more precise measures of the ECI of regions and ICI of industries, as noise and size effects are eliminated. The iterations are stopped when the ranking of regions and industries is stable from one step to another (i.e. no further information can be extracted from the structure of the region-industry network). The complexity indexes presented in this paper is based on N = 20 iterations[3].

Figure 2 shows the Brazilian micro-regions, which are classified according to the average of the ECIs in the period between 2006 and 2016. The figure shows that the distribution of complexity in Brazil is very concentrated, mainly in the microregions of São Paulo. Table 1 shows the averages of the ICIs that are grouped by division level of CNAE 2.0 during the period between 2006 and 2016. The table shows that the most complex industrial sectors are generally related to computer and electronic equipment, transportation, machinery and equipment, electrical material and chemicals during the period between 2006-2016.

FIGURE 2: Average ECI of the Brazilian microregions in the period 2006-2016.



Source: Authors' elaboration based on data from RAIS.

[3] For more details, see Hidalgo and Hausmann (2009).

TABLE 1: Average of the industry complexity index (ICI) according to the level of division of CNAE 2.0 - 2006 to 2016.

Sectors	Section of CNAE 2.0	Section description	Division of CNAE 2.0	Division description	ICI average	Number of microregions
Agriculture/Animal Farming and Extractive Industries	B	Extractive Industries	06	Extraction of oil and natural gas	0.193	16
	B	Extractive Industries	09	Mining support activities	-0.268	52
	A	Agriculture/Animal Farming	03	Fishing and Aquaculture	-0.380	38
	B	Extractive Industries	05	Coal extraction	-0.658	29
	B	Extractive Industries	07	Extraction of metal minerals	-0.664	17
Manufacturing	C	Processing Industries	26	Electronic Products	1.389	53
	C	Processing Industries	29	Motor Vehicles	1.244	82
	C	Processing Industries	28	Machinery and Equipment	1.190	73
	C	Processing Industries	27	Electrical Products and Materials	1.154	61
	C	Processing Industries	20	Chemical Products	0.825	72
Productive and distributive services	K	Financial Activities	66	Financial Services	1.106	73
	J	Information and Communication	62	Information Technology Services	1.085	237
	K	Financial Activities	65	Insurance and Pension Funds	1.072	98
	N	Administrative Activities	80	Surveillance and Security	0.936	141
	H	Transport and Postal Services	51	Airway Transport	0.826	43

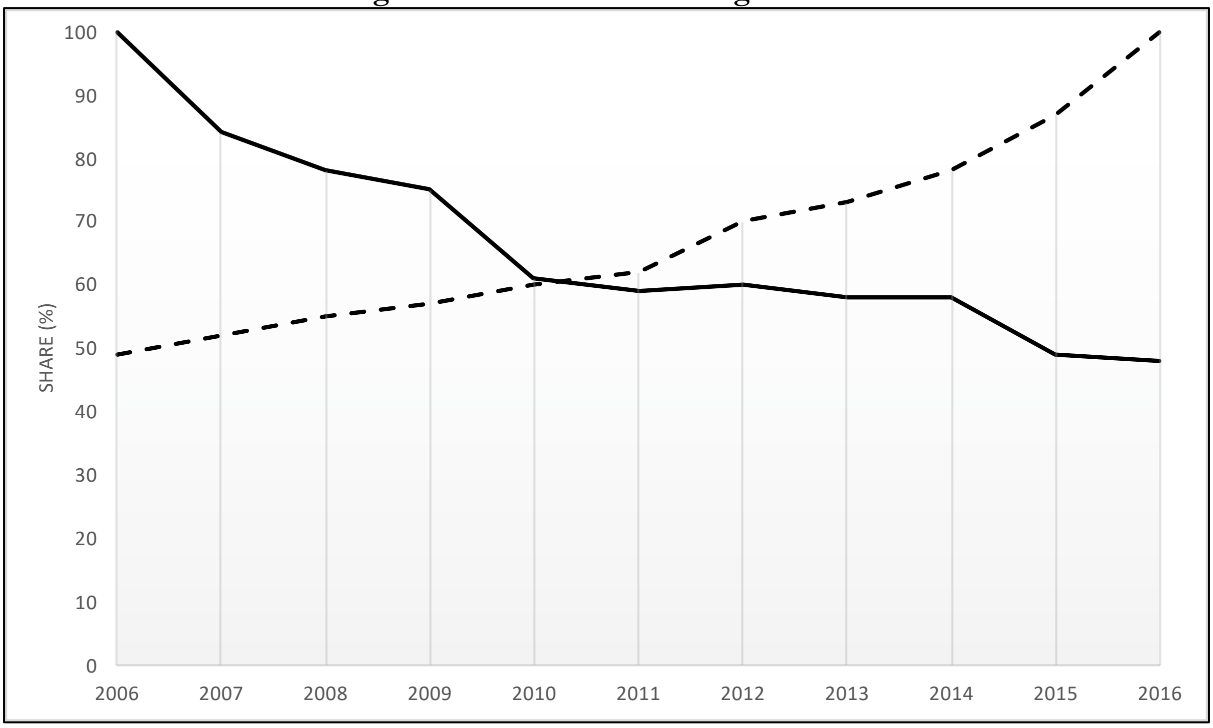
Source: Authors' elaboration based on data from RAIS.

3. RESULTS AND DISCUSSION

3.1 ENTRY, EXIT AND MAINTENANCE OF INDUSTRIES IN BRAZILIAN MICROREGIONS

A background of what we are trying to investigate is also whether the process of structural change is affected by the industrial relationship at the regional level. Figure 3 shows the change in the industrial composition of the Brazilian microregions throughout the study period. From 2006 to 2016, the Brazilian microregions underwent substantial structural changes. As far as local industries are concerned, only 48% of the industries that were specialized in 2006 were still specialized in 2016. Or, in other words, about half of the specialized industries that were present in 2006 disappeared from the productive structures of the microregions in 2016. From the reverse perspective, 49% of all local industries in 2016 already existed in 2006. These values are similar to those observed by Neffke *et al.* (2011) and Essletzbichlera (2015).

FIGURE 3: Structural change in the Brazilian microregions between 2006 and 2016.



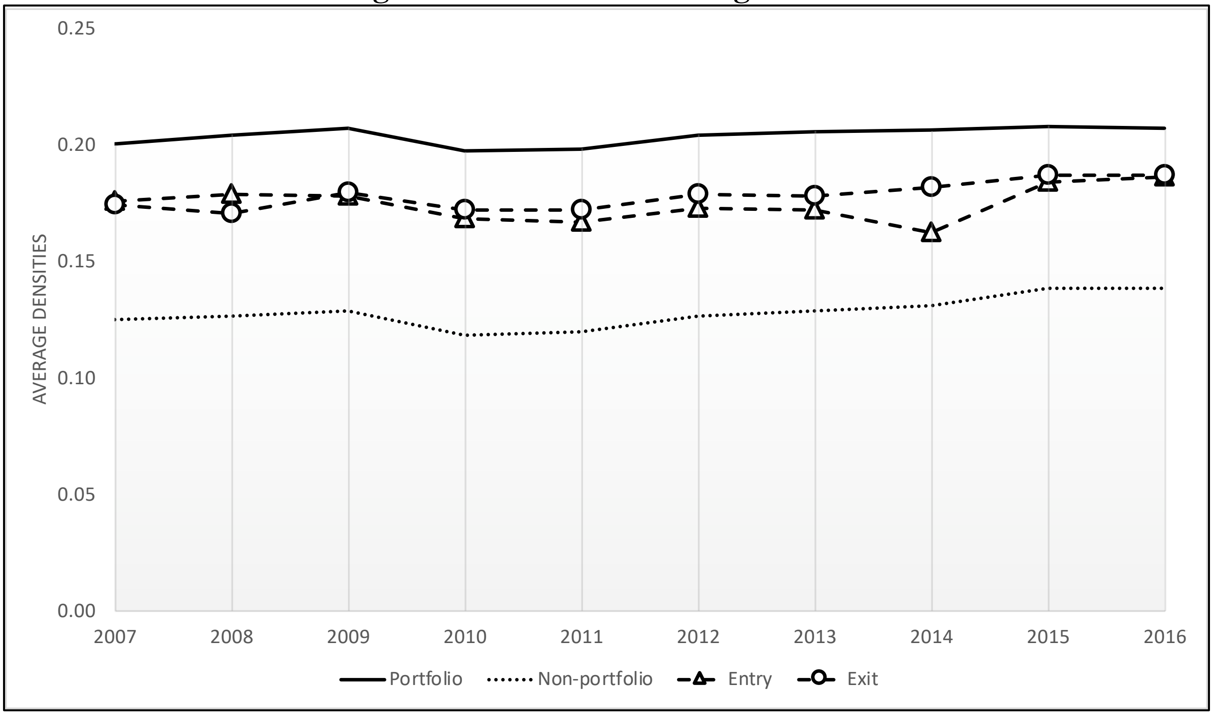
Source: Authors' elaboration based on Neffke *et al.* (2011) and Essletzbichlera (2015).
Note: The solid line represents the share of the number of specialized industries in the microregions that were present in these areas during 2006 and during each subsequent year. The dashed line shows the inverse perspective, that is, the share of the number of specialized industries in the microregions that were present in each of the previous years and that would still exist in 2016.

Figure 4 describes the evolution of industrial cohesion for the productive structure of all the Brazilian microregions between 2007 and 2016. The solid line brings the average of the densities of the industries that belong to the portfolio of the microregions each year. According to Neffke *et al.* (2011), a regional structure is considered cohesive if the average density of the industries that belong to the region is greater than that of the industries that are not part of it (i.e. regions are considered cohesive if the solid line is above the dotted line). According to Figure 4, the cohesion of micro-regions is on average stable over time.

It is also useful to examine how the entry and exit of industries influences the industrial cohesion of a micro-region. In Figure 4 the industries that entered were defined as those that were not specialized in a microregion (LQ <1) in a given year but became specialized (LQ > 1) the following year. On the other hand, the industries that came out were defined as those that were specialized in a microregion (LQ > 1) during a given year but did not become specialized (LQ <1) the following year.

The dashed line with the upward triangles denotes the average density of the industries that have entered the microregions each year. While the dashed line with circles denotes the average density of industries that have left the microregions each year, the line that represents the entries is always above the non-portfolio (dotted) line, which means that the industries that enter a region are much closer to the productive structure of the region than the industries that remained outside that structure. This suggested that the regions diversified towards sectors related to the existing industrial base.

FIGURE 4: Structural change in the Brazilian microregions between 2007 and 2016.



Source: Authors' elaboration based on Neffke *et al.* (2011) and Essletzbichlera (2015).
Note: The solid line brings the average of the densities of the industries that belong to the portfolio of the microregions each year. The dotted line brings the average of the densities of the industries that belong to the non-portfolio of the microregions each year. The dashed line with the upward triangles denotes the average density of the industries that have entered the microregions each year. While the dashed line with circles denotes the average density of industries that have left the microregions each year.

We also noticed that the line representing the outputs is above the dotted line (non-portfolio). This means that the industries that formerly belonged to the productive structure of the microregions were not completely disconnected from other economic activities in those regions. However, the output line is always well below the portfolio line (solid line). That is, although these existing industries were not completely alien to the other local industries, their position on the regional productive structure (on average) was less cohesive.

As the entry line is below the portfolio line, the entry weakens the industrial cohesion of the structures of the regions. Since the exit line is well below the portfolio line, the exit should increase the average cohesion of a region.

Overall, the results are similar to those found by Neffke *et al.* (2011) and by Essletzbichlera (2015), despite the fact that the relationship is measured differently, and the economic-geographical context differs for these two cases. More specifically, the three main results identified by these authors are confirmed in this analysis: first, the productive structures of the regions are cohesive and remain so over time. Second, industries are more likely to enter a region if they are related to the existing regional structure of industries. Third, industries that are less attached to the regional structure than other members of the structure are more likely to leave the region. These three findings are examined in more detail below.

3.2. SPECIFICATION OF THE ECONOMETRIC MODEL

Neffke *et al.* (2011) performed a similar test to that proposed in this article for Sweden in which dummy variables were defined for the entry, exit and maintenance of firms in the industrial structure of a region. We will follow the suggestion of these authors, however, defining the maintenance dummy with an assumed value of 1 if a microregion m is specialized in an economic activity s at the time t was also specialized at time $t + 5$, that is, if a microregion m possessed $LQ>$ 1 in period t in period $t + 5$. The entry dummy assumes the value 1 if a micro-region m was not specialized in an economic activity s at time t but was specialized at time $t + 5$. The dummy exit assumes the value 1 if a microregion m was specialized in an economic activity s at time t but was not specialized at time $t + 5$. Formally:

$$Maintenance_{m,s,t+5} = I(s \in P(m,t) \ \cap s \in P(m,t+5)) \tag{25}$$

$$Entry_{m,s,t+5} = I(s \in P(m,t) \ \cap s \notin P(m,t+5)) \tag{26}$$

$$Exit_{m,s,t+5} = I(s \notin P(m,t) \ \cap s \in P(m,t+5)) \tag{27}$$

We want to estimate how the relationship between an economic activity and the structure of the Brazilian microregions influences the maintenance, entry and exit of sectors of the microregions. This relationship as we describe is captured by the density variable. The basic econometric equation to be estimated can be written as follows:

$$\begin{aligned} Y_{m,s,t+5} = \alpha_1 + \alpha_2 Density_{m,s,t} + \alpha_3 ECI_{m,t} + \alpha_4 ICI_{s,t} + \alpha_4 LQ_{m,s,t} \\ + \alpha_5 Diversity_{m,t} + \alpha_6 \ln(Population)_{m,t} + \phi_m + \psi_s + \lambda_t + \varepsilon_{m,s,t} \end{aligned} \tag{28}$$

in which the dependent variable $Y_{m,s,t+5}$ represents the dummy variables of maintenance, entry and exit of an economic activity s in a micro-region m in period $t + 5$.

Following the theoretical reference, our main variables of interest are Density, which indicates how the economic activity is related to the pre-existing set of capabilities of a micro-region, and the economic complexities of the microregions (ECI) and the industries (ICI), which evaluate the modernization of the productive structure of the micro-region.

In order to capture the effects of the economic structure that impact the agglomeration forces, we follow the literature (Glaeser *et al.*, 1992; Ciccone and Hall, 1996; Combes, 2000; Glaeser and Maré, 2001; Combes *et al.*, 2008) with use of representative indicators of specialization and diversity of economic activities and the size of the microregions.

The locational quotient (*LQ*) will be the proxy measure for industrial specialization, the source of location externalities, is described according to Equation 7. The diversity indicator used is represented by the Shannon index[4], formally described as:

$$Diversity_{m,t} = e^{H} \tag{29}$$

such that $H = -\sum_{s=1}^{S} p_s \ln p_s$ and $p_s = {}^{emp_{m,s,t}}/_{emp_{m,t}}$.

In what: $emp_{m,s,t}$ is the employment of sector s in the microregion m in period t; $emp_{m,t}$ is the total employment in region m in period t.

In addition, the base econometric model used, defined by Equation 28, is a three-way fixed-effects model that takes into account the possible variable biases omitted at state, sector, and time levels: ϕ_m represent the fixed effects estimated directly by inclusion of dummy variables for each state; ψ_s represent the fixed effects estimated directly by the inclusion of dummy variables for each division of CNAE 2.0; λ_t represent the fixed effects estimated directly by the inclusion of dummy variables for each year. In addition, $\varepsilon_{m,s,t}$ is an error term i.i.d for other influences not observed.

Our panel consists of data from 568 Brazilian microregions, 1162 economic activities at the subclass level of CNAE 2.0 (see Table A.1 in Appendix) for two five-year periods between 2006 and 2016 (2006-2011 and 2011-2016) denoting the first year of these two periods for the last year of $t + 5$. The result was a balanced panel with 1,295,676 observations. Table 2 provides some summary statistics of the variables used in econometric analysis.

[4] See Shannon (1948).

TABLE 2: Descriptive statistics of the variables used in the econometric analysis.

Variables	Number of observations	Average	Standard deviation	Min	Max	Correlation					
						Density	ECI	ICI	LQ	Diversity	Population
Maintenance	1,295,676	0.078	0.268	0	1						
Entry	1,132,800	0.060	0.237	0	1						
Exit	162,876	0.379	0.485	0	1						
Density	1,255,500	0.131	0.070	0.003	0.648	1.000					
ECI	1,295,676	0.000	0.999	-1.857	4.116	0.816	1.000				
ICI	1,295,676	0.000	1.000	-2.206	4.525	-0.216	0.000	1.000			
LQ	1,275,030	1.205	40.869	0.000	30,432.660	0.009	-0.003	-0.021	1.000		
Diversity	1,295,676	68.502	46.094	2.835	270.643	0.898	0.859	-0.001	-0.003	1.000	
Population	1,295,676	339,740	875,832	2,321	13,882,809	0.743	0.735	0.000	-0.003	0.767	1.000

Source: Authors' elaboration based on data from RAIS.

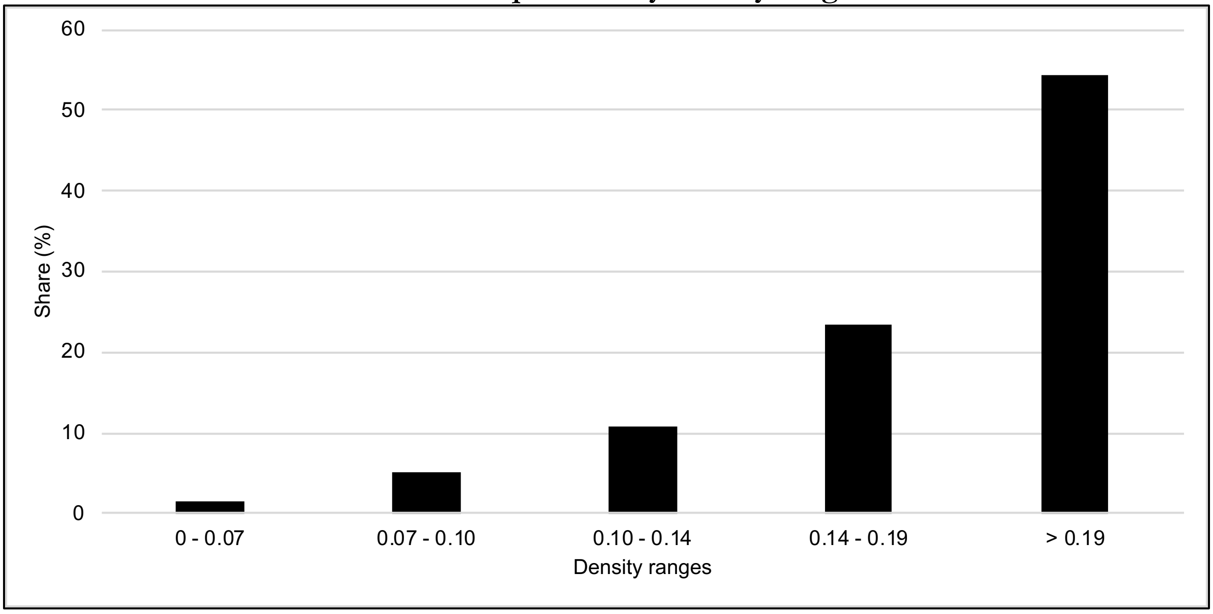
For calculations involving the maintenance dummy variable, all combinations of economic activities and micro-regions were considered as observations (total of 1,295,676 observations). However, if an industry is already present in a micro-region it obviously cannot enter the region industries. Reasoning in an analogous way if an industry is absent in the micro-region, then it is obviously impossible to leave the region. Therefore, in our calculations that involve the entry dummy variable, we used the subsample of industries that were absent from the region in year t (total of 1,132,800 observations), that is, we considered those industries that would be candidates to enter the microregions. On the other hand, in the calculations involving the exit dummy variable, we used the subsample of industries that were present in the region in year t (total of 162,876 observations), that is, they would be candidates to leave the microregions.

In order to determine the economic importance of proximity to regional productive structures, we analyzed how density affects the entry, exit and maintenance probabilities of industries. In the sample, during each of the years, there were 67,629 events in which an industry entered a region., that is, an industry came to have $LQ > 1$ in a microregion. An industry could only entry a given microregion in a given period if it did not yet belong to the productive structure of the region at the beginning of the period, that is, if $LQ < 1$ in year *t*. In total, given that there were 1,132,800 entry opportunities, we estimate the average entry probability to be 67,629 / 1,132,800 = 5.9%. Likewise, the average exit probability was 37.9% because there were 61,736 events of an industry leaving one region and 162,876 exit opportunities. Finally, we estimate that the average maintenance probability is 8.0% (out of a total of 1,295,676 possible regional industries, 101,140 already existed in year *t*).

For each sample, density values were grouped into density ranges with defined intervals according to the division of each sample into five quintiles of distribution. Figure 5, 6 and 7 show, respectively, these relative frequencies as a percentage of the maintenance, entry and exit of economic activities of the regional productive structure by density ranges. These results reveal how these percentages change as the density increases.

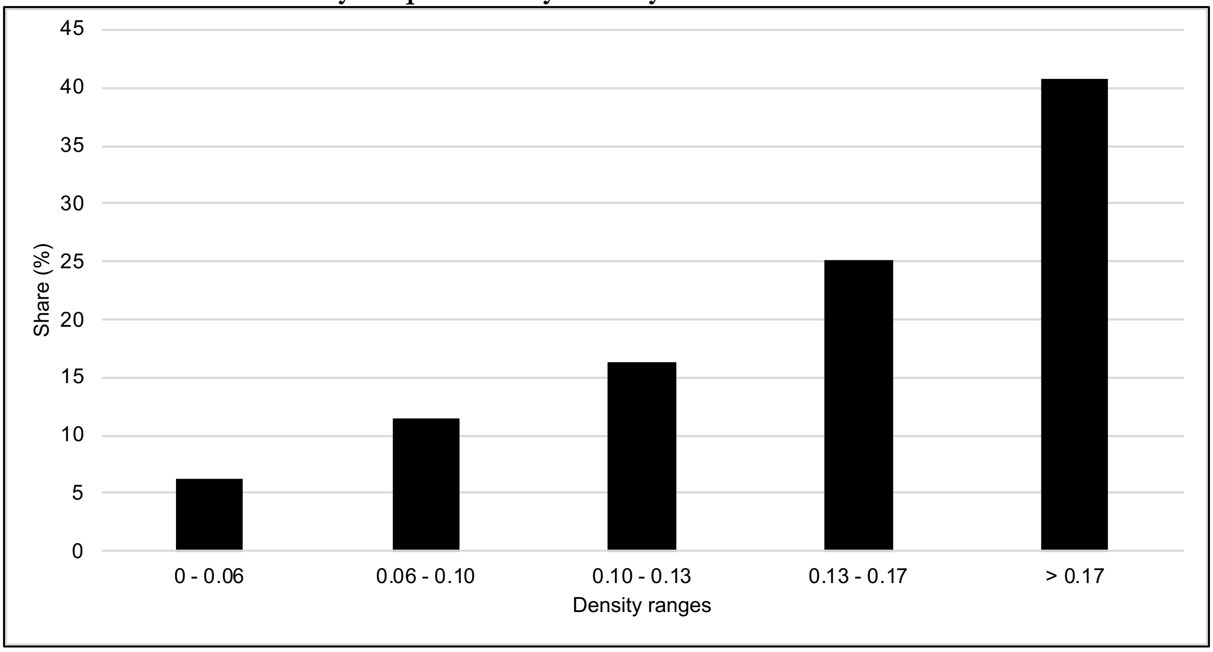
As you move along the horizontal axis of densities, you can see that maintenance and entry holdings are, at the beginning, below their general averages, but end well above these averages. The percentage of maintenance events is 1.4% in the first density range, and 54.4% in the last range. On the other hand, the percentage of entry events increased from 6.3% in the first group to 40.7% in the last group. Conversely, the percentage of exit events starts at 21.3%, but decreases to 11.2% in the last range.

FIGURE 5: Relative maintenance frequencies by density ranges.



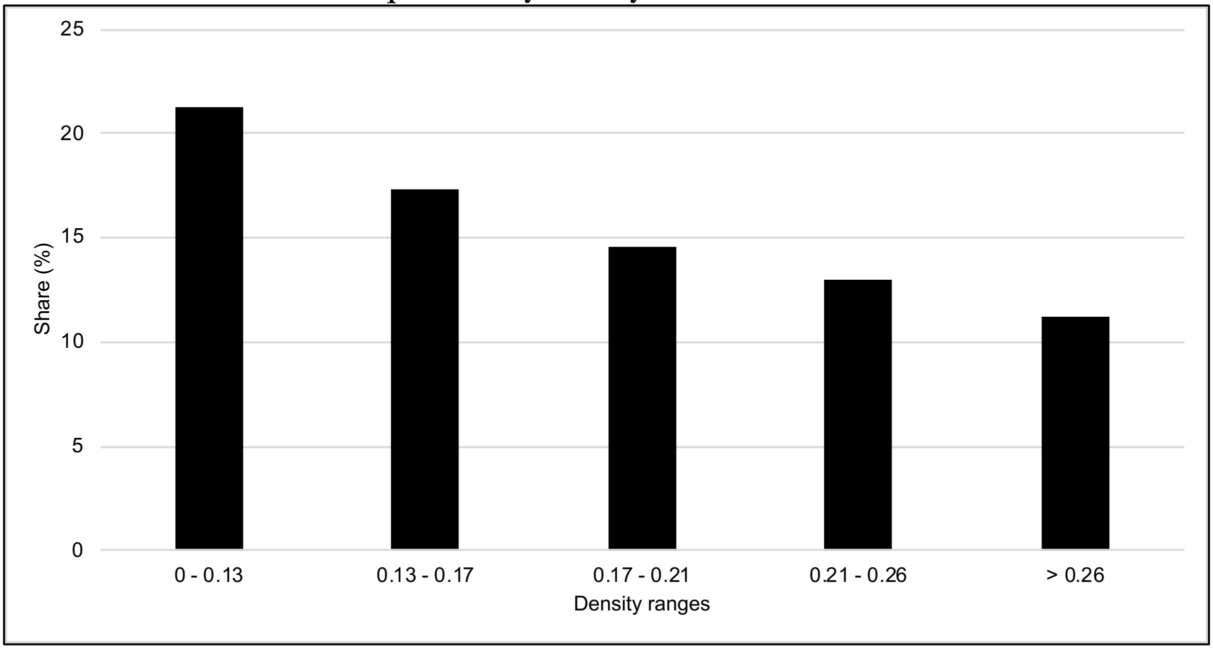
Source: Authors' elaboration based on data from RAIS.

FIGURE 6: Relative entry frequencies by density bands.



Source: Authors' elaboration based on data from RAIS.

FIGURE 7: Relative exit frequencies by density bands.



Source: Authors' elaboration based on data from RAIS.

After presenting our identification strategy, the variables we will use, and our database, we will present the results of the estimates (see following section). Estimates were made for all Brazilian microregions for two five-year periods between 2006 and 2016 (2006-2011 and 2011-2016).

3.3. ECONOMETRIC RESULTS

This section presents the results of the econometric model described in Equation 28 that analyzes: i. the probability that a micro-region will become specialized in a new economic activity that is related to its productive structure; ii. the probability that a micro-region will no longer be specialized in an economic activity if it is not related to its productive structure; iii. the probability that a micro-region will become specialized in a new economic activity that is complex; iv. the probability that a complex microregion will become specialized in a new economic activity that is complex.

We estimated Equation 28 using Logit model. All the regressions performed corrected for heteroskedasticity by the robust standard error procedure. The following Table 3 present the results of the regression analysis with binary maintenance variables (columns I to IV), as well as the entry (columns V to VIII) and exit (columns IX to XII) as dependent variables. First, we present the results for a simpler equation containing only the density and complexity variables, treating the model with fixed effects. Next, we present our complete model proposed by Equation 28, and we also consider a version of the model without fixed effects.

Therefore, we expect a positive coefficient for density in the maintenance and entry models (hypothesis 1), a negative coefficient for density in the exit models (hypothesis 2), a negative coefficient for the complexity variables in the maintenance and entry (hypothesis 3).

First, we will analyze the results of the estimates for the variables used as controls for externalities. The coefficient of the variable that indicates the specialization (*LQ*) presents a positive result and is statistically significant in the estimates using the maintenance and entry variables as dependent variables. However, the results for the dependent variable exit, in general, revealed negative, but statistically insignificant signs.

In relation to the variables *Diversity* and *ln(Population)*, which depict the size and diversification of the productive structure of the microregion, they presented negative results and were statistically significant in the estimates using the variables of maintenance and entry as dependents. On the other hand, the results for the exit dependent variable presented a positive and statistically significant sign.

These results indicate that the more specialized the microregion, the more cohesive the productive structure tends to increase the probability of entry of related sectors. On the other hand, the larger and more diversified the microregion, the less cohesive the productive structure tends to be and the lower the likelihood of entry of related sectors. The results also indicate that the specialization of the microregion tends not to affect the probability for the exit of sectors. On the other hand, size and diversity tend to reduce the relatedness of the microregion by increasing the probability of leaving sectors.

The main objective of this manuscript is to test the empirical validity of the related regional diversification hypothesis. The presented results reveal that the process of regional diversification is conditioned by existing regional structures. In this sense, we now evaluate the results of the estimates of the coefficients of the variable *Density* in the models of maintenance, entry and exit of economic activities in a micro-region.

The signs of all estimates were as expected and are statistically significant. The proximity of an industry to the regional productive structure (measured by density) increases the likelihood that the industry remains a member of this regional structure, or, if it is not yet a member, that likelihood increases that the industry will enter the region within a five-year period. On the other hand, the negative coefficient (columns IX to XII) the probability of an industry leaving a region increases if it is not related to the sectors of the micro-region.

To summarize, regression analyses confirm that the three regularities we identified in the previous section are accurate. The proximity of an industry to a regional structure has important consequences for the industrial cohesion of a micro-region and for the evolution of its industrial structure.

As expected, the regions are less likely to develop new specializations in complex technological activities (hypothesis 3). The effect of the economic complexity of the microregion (ECI) on maintenance is positive and significant when we added regional controls and fixed effects (columns III and IV). Overall, when the complexity of a micro-region increases the relative probability that a region remains an expert in a given economic activity increases. Regarding ECI in the entry models, the ECI coefficient signal is negative and significant (only in column VIII the result of this coefficient was not significant). That is, when the complexity of a micro-region increases, the relative probability of a region to become an expert in a new sector decrease.

TABLE 3: Results of the estimated models.

	Maintenance				Entry				Exit			
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)	(XII)
Density	22.7698***	24.886***	38.9011***	46.7479***	12.7790***	12.7020***	16.4660***	19.9691***	-8.0594***	-8.7009***	-11.4426***	-15.7051***
	(0.0967)	(0.1167)	(0.2715)	(0.4551)	(0.1164)	(0.1312)	(0.1748)	(0.2096)	(0.1363)	(0.1561)	(0.2119)	(0.2564)
ECI	-0.7982***	-0.9993***	0.0375***	0.1363***	-0.3882***	-0.3799***	-0.0950***	-0.0064	0.2255***	0.3056***	-0.0841***	-0.0783***
	(0.0071)	(0.0087)	(0.0104)	(0.0123)	(0.0082)	(0.0095)	(0.0095)	(0.0113)	(0.0100)	(0.0122)	(0.0119)	(0.0149)
ICI	-0.6614***	-0.7784***	-0.4672***	-0.5587***	-0.3060***	-0.3947***	-0.0720***	-0.1289***	0.2767***	0.3302***	0.3324***	0.3858***
	(0.0050)	(0.0068)	(0.0071)	(0.0089)	(0.0049)	(0.0071)	(0.0053)	(0.0074)	(0.0088)	(0.0116)	(0.0095)	(0.0127)
LQ			0.0866***	0.0799***			2.2971***	2.0294***			-0.0011	-0.0018*
			(0.0212)	(0.0186)			(0.0163)	(0.0186)			(0.0007)	(0.0010)
Diversity			-0.0319***	-0.0401***			-0.0150***	-0.0174***			0.0088***	0.0142***
			(0.0004)	(0.0005)			(0.0003)	(0.0003)			(0.0003)	(0.0004)
ln(Population)			-0.4354***	-0.7541***			-0.1140***	-0.2880***			0.1851***	0.2311***
			(0.0078)	(0.0122)			(0.0077)	(0.0094)			(0.0106)	(0.0154)
Constant	-6.2536***	-7.9619***	-1.3044***	-0.7273***	-4.6509***	-5.3056***	-2.9950***	-1.9092***	0.9248***	1.4989***	-1.4233***	-0.6866***
	(0.0157)	(0.0398)	(0.0968)	(0.1455)	(0.0169)	(0.0432)	(0.0874)	(0.1097)	(0.0230)	(0.0610)	(0.1204)	(0.1779)
F.E. States	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
F.E. Industries	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
F.E. Year	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Pseudo R2	0.2334	0.2684	0.3375	0.3826	0.0677	0.1051	0.1227	0.1459	0.0370	0.0791	0.0449	0.0757
Wald chi2	133,293.95	137,648.74	127,137.14	129,207.41	37,360.71	51,437.00	67,246.81	75,236.80	6,094.05	12,477.22	7,054.95	11,173.79
Number of observations	1,255,500	1,255,500	1,251,594	1,251,594	1,111,705	1,111,705	1,111,705	1,111,705	143,795	143,795	139,889	139,889

Source: Authors' elaboration based on data from RAIS, initial periods (t) are 2006 and 2011 and final ($t + 5$) 2011 and 2016.
Note: The robust standard errors of each estimate are enclosed in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. The year 2006 was chosen as the start year in order to avoid any remaining inconsistencies of changes in industry rankings between 2005 and 2006.

Regarding the effect of the economic complexity of the industry (ICI), the sign of the ICI coefficient was negative and significant in all models when we estimated the variables maintenance and entry as dependent. On the other hand, it presented a positive and significant signal in the models that consider the dependent variable exit. These results reflect a dilemma of diversification brought about by the low complexity trap. A complex activity is more attractive, but at the same time it is also more difficult to produce (negative effect on entry and maintenance and positive on exit). Therefore, the relationship between complexity and the new specialization is not linear and may be region specific. Hypothesis 4 indicates that complex regions are more likely to develop new specializations in complex technological activities. To investigate this, we divided the sample to those observations with a high level of economic complexity of the microregion. We did this from the 4th quartile of the microregion sample for each year according to ECI values. The results are shown in Table 4.

TABLE 4: Estimated results for a sample containing the fourth quartile of the observations defined from ECI.

	Maintenance	Entry	Exit
Density	41.4964***	18.903***	-17.7706***
	(0.4527)	(0.3213)	(0.5089)
ECI	0.3635***	0.1485***	-0.2046***
	(0.0187)	(0.0187)	(0.0345)
ICI	-0.2086***	0.1627***	0.1514***
	(0.0118)	(0.0107)	(0.0207)
LQ	0.5279***	2.3113***	-0.0348***
	(0.0331)	(0.0272)	(0.0053)
Diversity	-0.0311***	-0.0132***	0.0133***
	(0.0005)	(0.0004)	(0.0007)
ln(Population)	-1.1428***	-0.6684***	0.4986***
	(0.0229)	(0.0201)	(0.0352)
Constant	3.9594***	2.8726***	-4.6972***
	(0.2752)	(0.2400)	(0.4136)
F.E. States	Yes	Yes	Yes
F.E. Industries	Yes	Yes	Yes
F.E. Year	Yes	Yes	Yes
Pseudo R2	0.4169	0.1428	0.0721
Wald chi2	30,254.51	21,958.17	2,481.88
Number of observations	311,777	276,917	36,746

Source: Authors' elaboration based on data from RAIS, initial periods (t) are 2006 and 2011 and final (t + 5) 2011 and 2016.
Note: The robust standard errors of each estimate are enclosed in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. The year 2006 was chosen as the start year in order to avoid any remaining inconsistencies of changes in industry rankings between 2005 and 2006.

The main result is that the economic complexity of microregions, which is measured by ECI, influences the entry of new complex economic activities in the microregion. When the ECI level is high, i.e. when the models include only 25% microregions with the highest ECIs, the economic complexity of the industry (measured by the ICI) has a positive and significant impact on the specialization of new economic activities. Another interesting result is that even in the most complex microregions in the country, when we use exit as a dependent variable the complexity of the sector (ICI) has a positive and significant impact and when we use maintenance as a dependent variable (ICI) in general, has a negative and significant impact. These results reveal that the complexity of the microregion conditions access to new complex industries, however, does not completely solve the trap of low complexity because the more complex sectors are less likely to remain specialized even in the regions that are more complex. This confirms (in part) hypothesis 4.

4. CONCLUSIONS

This paper complements and expands the conceptual and empirical literature on relatedness and regional evolution as a process of industrial diversification (Frenken *et al.*, 2007; Boschma and Iammarino, 2009; Neffke *et al.*, 2011; Boschma *et al.*, 2013; Essletzbichlera, 2015). We find evidence that the productive specialization of a Brazilian micro-region in a new economic activity is conditioned by the productive structure that exists in the microregions, as a process strongly dependent on trajectory. The regions diversify, branching into sectors related to their current sectors. Analyzing the evolution in the productive structures of 568 Brazilian microregions during the period between 2006 and 2016, the results indicated that a new industry is more likely to enter a micro-region when technologically related to other industries in that region. Also, an existing industry was more likely to leave a micro-region when it was not related to other industries in that region.

However, in general, industries entering a region are less related to the local industrial portfolio than the average relationship between the members of the existing productive structure. Consequently, entry reduces the technological cohesion of a region by adding new variety. Output probabilities, by contrast, increase as industries hold positions more technologically distant from the productive structure of a region. Thus, the output increases the technological cohesion of the regions.

These analyses suggest that it is difficult to attract new industries to a region if they are technologically distant from current local activities. Moreover, even if they enter, the exit probabilities are high if they are technologically distant from local activities. This difficulty becomes even greater in the case of complex industries. We picture this as a diversification dilemma brought about by the trap of low complexity.

On the other hand, for the industries that are related to the current productive structure, but which have not yet entered the local economy, the region may be poorly adapted for reasons that are not investigated in this manuscript. The results point to a number of future research questions with important policy implications.

It would therefore be interesting to investigate these potential entrants more closely and determine why companies in these industries seemingly avoid the region. If bottlenecks can be remedied, policy initiatives aimed at removing bottlenecks will be of great value.

In addition, it is important to examine how the technological cohesion of regions is linked to their performance, changes in employment rates and unemployment, productivity and growth. Do regions that are more technologically cohesive and perform better than those that are not?

REFERENCES

Almeida, P.; Kogut, B (1999) Localization of knowledge and the mobility of engineers in regional networks, **Management Science**, v.45, n.7 p.905-917.

Balassa, B. (1965) Trade Liberalization and Revealed Comparative Advantage. **Manchester School of Economic and Social Studies**, v.33, p.99-123.

Balland, P.A.; Rigby, D. (2017) The Geography of Complex Knowledge, **Economic Geography**, v.93, n.1, p.1-23.

Balland, P.A., Boschma, R., Crespo, J., Rigby, D.L. (2018) Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. **Regional Studies**, p.1-17.

Brezis, E.S; Krugman, P. (1997) Technology and the Life Cycle of Cities, **Journal of Economic Growth**, v.2, p.369-383.

Boschma, R. (2014) Constructing regional advantage and smart specialization: Comparisons of two European policy concepts, **Italian Journal of Regional Science**, v.13. n.1, p.51-68.

Boschma, R., Minondo, A.; Navarro, M. (2013) New industries and relatedness in regions. **Economic Geography**, v.89, p.29-51.

Boschma, R.; Frenken, K. (2009) Technological relatedness and regional branching. **Papers in Evolutionary Economic Geography**, n.0907.

Boschma, R.; Frenken, K. (2011) Technological relatedness, related variety and economic geography. In: COOKE, P.; ASHEIM, B.; BOSCHMA, R.; MARTIN, R.; SCHWARTZ, D.; TODTLING, F., **The Handbook on Regional Innovation and Growth**. Cheltenham. Edward Elgar.

Boschma, R..; Iammarino S. (2009) Related variety, trade linkages and regional growth in Italy, **Economic Geography**, v.85, n.3, p.289-311.

Bryce, D.J.; Winter, S.G. (2009) A General Interindustry Relatedness Index. **Management Science**. v.55, n.9, p.1570-1585.

Ciccone, A., Hall, R.E. (1996) Productivity and the density of economic activity. **American Economic Review**, v.86, n.1, p.54-70.

Combes, P.P. (2000) Economic structure and local growth: France, 1984-1993. **Journal of Urban Economics**, v.47, p.329-355.

Combes, P.P.; Duranton, T.; Gobillon, J.M.. (2008) Spatial wage disparities: Sorting matters! **Journal of Urban Economics**. v.63, n.2, p.723–742.

Crocco, M.A.; Galinari, R.; Santos, F.; Lemos, M.B.; Simões, R. (2006) Metodologia de identificação de aglomerações produtivas locais. **Nova Economia**. v.16, n.2, p.211-241.

Ellison, G.; Glaeser, E. L., (1997) Geographic concentration in U.S. manufacturing industries: A dartboard approach. **Journal of Political Economy**, v.105 n.5, 889-927.

Engelsman E.C.; Van Raan A.F.J. (1991) Mapping of Technology, A First Exploration of Knowledge Diffusion Amongst Fields of Technology. The Hague: **Policy Studies on Technology and Economy** (BTE) Series No. 15.

Engelsman E.C.; Van Raan, A.F.J. (1994), A patent-based cartography of technology. **Research Policy**, v.23, n.1, p.1-26.

Essletzbichler, J. (2015) Relatedness, regional branching and technological cohesion in US metropolitan areas. **Regional Studies**, v.49, n.5, p.752-766.

Fan, J.P.H.; Lang, L.H.P. (2000) The Measurement of Relatedness: An Application to Corporate Diversification. **Journal of Business**. v.73, n.4, p.629-660.

Farjoun, M. (1994) Beyond Industry Boundaries: Human Expertise, Diversification and Resource-Related Industry Groups. **Organization Science**. v.5, n.2, p.185-199.

Fleming, L.; Sorenson, O. (2001) Technology as a complex adaptive system: evidence from patent data. **Research Policy,** v.30, n.7, p.1019–1039.

Frenken, K.; Van Oort, F.G.; Verburg, T. (2007) Related variety, unrelated variety and regional economic growth, **Regional Studies**, v. 41, n.5, p.685-697.

Gilsing, V.; Nooteboom, B.; Vanhaverbeke, W.; Duysters, G.; Van Den Oord, A. (2007) Network embeddedness and the exploration of novel technologies. Technological distance, betweenness centrality and density, **Research Policy**, v.37, n.10, p.1717-1731.

Glaeser, E.L., Maré, D.C. (2001) Cities and skills. **Journal of Labor Economics**, v.19, n.2, 316-342.

Glaeser, E.L.; Kallal, H.D.; Scheinkman, J.A.; Schleifer A. (1992) Growth in Cities, **Journal of Political Economy**, v.100, n.6, p.1126-1152.

Greene, W.H. (2003) **Econometric analysis**. Upper Saddle River, N.J.: Prentice Hall.

Hausmann, R.; Hidalgo C.A.; Bustos, S.; Coscia, M.; Chung, S.; Jimenez, J.; Simões, A.; Yildirim, M. A (2011) **The Atlas of Economics Complexity – Mapping Paths to prosperity**. Puritan Press, p. 364.

Hausmann, R.; Hidalgo, C.A. (2010), Country diversification, product ubiquity, and economic divergence, **CID Working Paper N. 201**, Center for International Development, Harvard University.

Hausmann, R.; Hidalgo, C.A. (2011) The network structure of economic output. **Journal of Economic Growth**, v. 16, n. 4, p. 309-342.

Hausmann, R.; Klinger B. (2007) The structure of the product space and the evolution of comparative advantage. **CID working paper n. 146**, Center for International Development, Harvard University, Cambridge.

Henderson, J.V.; Kuncoro, A.; Turner, M. (1995) Industrial development in cities, **Journal of Political Economy**, v.103, n5, p.1067-1085.

Hidalgo, C.; Hausmann, R. (2009) The building blocks of economic complexity, **Proceedings of the National Academy of Sciences**, v.106, n.26, p.10570–10575.

Hidalgo, C.A.; Klinger, B.; Barabási, A.-L.; Hausmann, R. (2007) The Product Space Conditions the Development of Nations, **Science**, v.317, n.5837, p.482-487.

Jacobs, J. (1969) **The Economy of Cities**. New York: Vintage Books.

Kogut, B.; Zander, U. (1993) Knowledge of the firm and the evolutionary theory of the multinational corporation, **Journal of International Business Studies**, v.24, n.4, p.625-645.

Menzel, M. (2008) Dynamic Proximities – Changing Relations by Creating and Bridging Distances. **Papers in Evolutionary Economic Geography**, n.0816.

Neffke F.M.; Henning, M.S. (2008) Revealed Relatedness: Mapping Industry Space. **DRUID Working Paper**, n.08-18.

Neffke, F.M.; Henning, M.S; Boschma, R. (2011) How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. **Economic Geography**, v.87, n.3, p.237-65.

Nooteboom, B. (2000) **Learning and innovation in organizations and economies**. Oxford: Oxford University Press.

Penrose, E. (1959) **The Theory of the Growth of the Firm**. Oxford: Oxford University Press.

Porter, M. E. (2003) The economic performance of regions. **Regional Studies**, v.37, p.549–78.

Rigby, D. (2013) Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes, **Regional Studies**, v.49, n.11, p.1922-1937

Scherer, F.M. (1982) Inter-Industry Technology Flows in the United States. **Research Policy**. V.11, p.227-245.

Shannon, C.E. (1948) A mathematical theory of communication. **Bell System Technical Journal**. v.27, p.379-423

Simões, R.; Hermeto, A.; Amaral, P. (2006) Rede urbana metropolitana: uma análise da estrutura terciária de Belo Horizonte. **Ensaios FEE**, v.27, n.2, p.471-513.

Storper, M. (1995) The resurgence of regional economies, ten years later: The region as a nexus of untraded interdependencies. **European Urban and Regional Studies** v.2, p.191–221.

Teece D.J.; Rumelt R.; Dosi G.; Winter S.G. (1994) Understanding corporate coherence. Theory and evidence. **Journal of Economic Behaviour and Organization**. v.23, p.1-30.

APPENDIX

TABLE A.1: Selected sectors of CNAE 2.0 by description and codes.

Section	Divisions	CNAE description	Selected sector
A	01 .. 03	Agriculture and Animal Farming	Yes
B	05 .. 09	Extractive Industries	Yes
C	10 .. 33	Processing Industries	Yes
D	35 .. 35	Electricity and Gas	Yes
E	36 .. 39	Basic Sanitation	Yes
F	41 .. 43	Construction	Yes
G	45 .. 47	Trade	Yes
H	49 .. 53	Transport and Postal Services	Yes
I	55 .. 56	Lodging and Food	No
J	58 .. 63	Information and Communication	Yes
K	64 .. 66	Financial Activities	Yes
L	68 .. 68	Real Estate Activities	Yes
M	69 .. 75	Specialized Services	Yes
N	77 .. 82	Administrative Activities	Yes
O	84 .. 84	Public Administration	No
P	85 .. 85	Education	No
Q	86 .. 88	Human Healthcare and Social Services	No
R	90 .. 93	Arts, Culture and Recreation	No
S	94 .. 96	Other Services	No
T	97 .. 97	Domestic Services	No
U	99 .. 99	International Organizations	No

Source: Authors' elaboration based on Simões, Oliveira and Amaral (2006)