Regional diversification in Brazil: the role of relatedness and complexity

Mariane Santos Françoso, Ron Boschma & Nicholas Vonortas

Papers in Evolutionary Economic Geography

22.06

Utrecht University Human Geography and Planning

Regional diversification in Brazil: the role of relatedness and complexity

Mariane Santos Françoso[1], Ron Boschma[2,3] and Nicholas Vonortas[4,5,6]

[1] InSySPo, Department of Science and Technology Policy, University of Campinas, Campinas-SP, Brazil

[2] Department of Human Geography and Planning, Utrecht University, Utrecht, The Netherlands

[3] UiS Business School, University of Stavanger, Stavanger, Norway

[4] Institute for International Science and Technology Policy & Department of Economics, The George Washington University, Washington-DC, USA

[5] São Paulo Excellence Chair, University of Campinas, Campinas-SP, Brazil

[6] Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, Moscow, Russian Federation

Contact Information:

Mariane Santos Françoso
Department of Science and Technology Policy, University of Campinas, Brazil
marianef@unicamp.br (corresponding author)
orcid: 0000-0002-9785-3469

Abstract

The paper contributes to the growing literature on the relationship between relatedness, complexity and regional diversification. It explores regional diversification in an emerging economy, focusing on diversification opportunities of regions with distinct levels of local capabilities. We investigate the importance of relatedness and economic complexity for sectoral and technological diversification in all regions of Brazil during the period 2006-2019. Regions tend to diversify in sectors/technologies requiring similar capabilities to those already available locally. In general, the higher the sector/technology complexity, the lower the probability of diversification. However, in high-complex regions, complexity reverses into a positive force for diversification. Our analysis shows catching-up and diversification prospects vary widely across different types of regions in Brazil.

Keywords: regional diversification; relatedness; complexity; emerging economies; Brazil

JEL codes: O25, O33, R11, O31

1. Introduction

A large body of literature has investigated the possibilities of countries in the Global South to catch-up (Keun, 2012; Petralia et al., 2017). Less attention has been devoted to catching-up of regions at the sub-national scale, despite the fact that many regions in the Global South have a strong ambition to move up the economic ladder (Yeung, 2021). Huge policy efforts are made to engage regions in new and more sophisticated activities that may leverage value capture. However, many regions fail to do so.

Failure to catch-up is often attributed to local factors, such as a lack of absorptive capacity, a weak research and educational infrastructure, or poor institutions. Scholars in evolutionary and complexity thinking (Boschma, 2017; Dosi, 1988; Hidalgo et al., 2007; Hidalgo et al., 2018; Keun, 2012) argue that economic diversification is a path-dependent process, in which regions tend to diversify into sectors, jobs, products and technologies related to their local capabilities. The rationale behind it is that related activities demand similar capabilities that are more easy and less costly to combine. Studies have demonstrated that related diversification in regions is indeed the most common pattern. Almost all of these studies have been conducted on the Global North (Essletzbichler, 2015; Neffke et al., 2011; Rigby, 2015). Few studies on regional diversification exist on the Global South (Alonso & Martín, 2019; Breul & Pruss, 2022).

Hidalgo and Hausmann (2009) argued it is crucial for countries to diversify into complex activities and to enhance the complexity of their economies. They found that economic complexity is positively correlated with country GDP. This is because complex activities represent sophisticated knowledge and require a wide diversity of capabilities that need to be combined. The less available these capabilities and knowledge are, the more exclusive will be the products those countries are able to produce. This exclusivity gives them an advantage over others, as they produce something not widely available while having capabilities that are not easy to transfer across space. Most studies investigating economic complexity focus on the national level (e.g. Hartmann et al., 2021). However, there is a rapidly expanding literature that analyzes the effect of complexity on growth and innovation at the regional scale, finding a positive relationship (e.g. Davies & Maré, 2021; Mewes & Broekel, 2020). These studies focus almost exclusively on regions in developed countries of the Global North.

What is more, there is still little understanding of the nature of diversification in distinct types of regions (Boschma, 2017). Studies tend to suggest that diversification opportunities can vary across different types of regions (Pinheiro et al., 2022; Xiao et al., 2018). This is expected to be especially true in countries in the Global South where regional inequalities are more pronounced. However, little research has been done on this topic with a particular focus on regions in the Global South, and how diversification opportunities of regions may differ in such context.

This paper has two objectives. First, we investigate the relative importance of relatedness and complexity on regional diversification in an emerging country. Based on two datasets on patents and sectors, we examine technological and industrial diversification in Brazil. As patents cover only few sectors of an economy, technological capabilities might be quite weak in many regions, especially in the Global South. Therefore, we also examine sectoral data which include all sectors in an economy, as industrial capabilities are expected to be more widespread across regions. Our study finds that relatedness is positively and complexity is negatively correlated with regional diversification. Complex activities are more difficult to develop by regions in general, with the exception of high-complex regions. Second, there is little known about diversification opportunities in different types of regions in the Global South. Brazil is an interesting case because spatial inequality is high: the country is home to regions with characteristics similar to those of developed countries as well as regions looking similar to those of underdeveloped countries (Galetti et al., 2021; Romero et al., 2021). Taking Brazil as a case, we investigate how regional diversification processes occur in an emerging economy context and how distinct regions may develop different patterns of diversification under the same national conditions.

The paper is organized as follows. In Section 2, we discuss the literature on relatedness, complexity and regional diversification. Section 3 provides details on data and the methodological approach. In Section 4, we present and discuss the empirical findings. Finally, we conclude the paper discussing the implications of the analysis.

2. Relatedness, complexity and regional diversification

The path-dependent character of economic change has been widely documented in evolutionary economics (Dosi, 1982; Nelson & Winter, 1982). Scholars have sought to explain economic growth considering the role of innovation, creative destruction and product variety (Aghion & Howitt, 1992; Romer, 1990). The catching-up literature (Abramovitz, 1986; Keun, 2012; Lall, 1992) has identified conditions that enable or hinder innovation in developing countries, such as technological and social capabilities

(Fagerberg & Srholec, 2018). However, little attention has been paid to diversification processes in regions in the Global South at the sub-national scale, in particular to the question whether the set of existing capabilities and knowledge conditions shapes regional diversification (Alonso & Martín, 2019).

The role of regional capabilities is still underestimated in the catching-up literature (Boschma, 2022b). It is a well-known fact that knowledge spillovers are often geographically bounded within countries (Audretsch & Feldman, 1996; Feldman & Kogler, 2010; Jaffe et al., 1993). Regions have a tendency to accumulate knowledge and specialize in specific activities over time, as knowledge production is often cumulative and localized (Dosi, 1982). Scholars have pointed to the importance of cognitive proximity for learning and innovation: local agents require absorptive capacity to exploit external knowledge (Boschma, 2004; Frenken et al., 2007; Nooteboom, 2011). Such proximity notions have been used to build a theoretical argument that innovation and diversification is a path-dependent process conditioned by local resources. Regions often diversify into sectors, jobs, products and technologies that are closely related to their local capabilities (Boschma, 2017; Hidalgo et al., 2018). This is because related activities require similar capabilities that are more easy and less costly to combine (Breschi et al., 2003; Teece, 1992). Hidalgo et al. (2007) applied this insight to explain why countries diversify into new export products that are close or proximate to those they already produce. Diversification is path- and place-dependent, not a random process.

Neffke et al. (2011) confirmed the overall importance of related diversification at the sub-national scale. This was followed by many other studies finding similar outcomes (e.g. Boschma et al., 2015; Colombelli et al., 2014; Essletzbichler, 2015; Rigby, 2015). The overwhelming majority of these studies focused on regions in the Global North. Few studies exist on regional diversification in the Global South. By and large, these studies confirm that related diversification is the rule in regions in countries like China, Vietnam, Mexico and Brazil. Zhu et al. (2017) and He et al. (2018) found a positive influence of relatedness on the development of new industries in Chinese regions. Breul and Pruß (2022) confirmed such a finding for regions in Vietnam. In the same vein, Alonso and Martín (2019) showed that relatedness had a positive effect on the development of comparative advantage in new export products in regions in Brazil and Mexico. Galetti et al. (2021) employed industry-skills co-occurrence to calculate industry relatedness and showed for Brazilian microregions that relatedness had a positive effect on industry entry and employment growth. Jara-Figueroa et al. (2018) showed how industry- and occupation-specific related knowledge of workers impact the survival of regional pioneers in Brazil. Their findings indicate that the growth and survival of pioneers are positively impacted by hiring workers with previous experience in related industries.

The inflow of external knowledge also matters, similarly to regions in the Global North (Andersson et al., 2013; Neffke et al., 2018). The inflow of knowledge through trade and foreign direct investment (FDI) not only had an impact on regional diversification but also had a tendency to weaken the influence of relatedness. He et al. (2018) argued that foreign investment and international trade can compensate for the lack of relatedness to some degree and foster the emergence of new industries in regions in China. Breul and Pruß (2022) showed that foreign-owned companies may act as global pipelines accessing external capabilities and contributing to diversification in unrelated industries in Vietnam. Alonso and Martín (2019) showed that capabilities coming from abroad, in the form of imports, had a positive effect on the development of new products in Mexico, but not in Brazil.

These studies already indicate that not all regions may be subject to related diversification in the same manner. While related diversification is still the rule, Petralia et al. (2017) and Pinheiro et al. (2021) found evidence that the importance of relatedness for diversification may vary according to the level of development of countries. They showed that relatedness has a higher impact on diversification in earlier development stages, while more developed countries are less impacted by relatedness and manage to make bigger jumps in their economic evolution now and then. At the sub-national level, Xiao et al. (2018) found that the influence of relatedness on industrial diversification depends on the innovation capacity of a region in Europe. Relatedness is more important in regions with a weak innovation capacity and less important in regions with a strong innovation capacity, as the latter enables regions to break from their past to a larger extent. This comes close to what Zhu et al. (2017) found for Chinese regions, showing that regions with high levels of FDI, R&D and human capital were able to diversify into less related industries. However, we still know little how diversification opportunities of regions vary between distinct types of regions (Boschma, 2017), especially in the Global South where regional inequalities are more pronounced.

Another key element in the diversification literature is to identify and assess opportunities of regions to move into more complex activities, as these activities are expected to bring higher economic benefits. This is especially important for developing countries that aim to move up the economic ladder or break out of a low-complexity trap

(Hartmann et al. 2021; Pinheiro et al. 2022). According to Hausmann and Hidalgo (2009), complex activities build on advanced knowledge and combine a variety of capabilities that are difficult to copy. The less available these capabilities and knowledge are, the more exclusive will be the activities regions are able to produce. This exclusivity gives them an economic advantage, as they produce something that is not widely available while having capabilities that are difficult to transfer across space. Hidalgo and Hausmann (2009) and Hausmann et al. (2014) showed that economic complexity is positively correlated with gross domestic product (GDP) levels of countries. More recently, studies also found a positive relationship between complexity, gross regional product (GRP) growth and innovation at the regional scale (Antonelli et al., 2020; Balland et al., 2020; Davies & Maré, 2021; Mewes & Broekel, 2020; Pintar & Scherngell, 2020). Rigby et al. (2022) found in Europe a positive relationship between economic growth of cities (in terms of GRP and employment) and their tendency to diversify into related and complex technologies.

So, regions have a strong economic incentive to develop new activities that make their economies more complex. However, not all regions may have the same capacity to do so. Studies show that complex activities are more geographically concentrated in the largest cities that offer high density and wide variety of capabilities (Balland et al., 2020; Balland & Rigby, 2017; Mewes & Broekel, 2020). Balland et al. (2019) found that regions in Europe have a hard time to diversify into complex activities unless they draw on related capabilities that are present in the region. Pinheiro et al. (2022) found that advanced regions in Europe with high levels of income and complexity are more capable of entering high-complex activities, while lagging regions rely more on low-complex activities when diversifying. However, these studies have focused on regions in developed countries in the Global North only.

Although some works on regional diversification deal with emerging economies and more specifically Brazil, none of them investigated how distinct regions, in terms of their complexity, show different diversification patterns. For instance, Galetti et al. (2021) found that relatedness is more important for diversification in lagging regions in Brazil, but they did not address how complexity levels of regions may impact their diversification patterns. Balland et al. (2019) proposed a framework to identify diversification opportunities of regions. They make use of the concepts of relatedness and complexity to assess the costs and benefits of all options each region might have. The relatedness concept is used to identify activities that might be developed at relatively low costs because they can build on relevant (related) capabilities in the region. The concept of complexity is employed to identify activities that may bring high economic benefits to the region. To our knowledge, no study yet exists that has explored diversification opportunities of regions in the Global South in terms of relatedness and complexity.

3. Data and methodology

3.1.Data

To investigate diversification opportunities of all Brazilian regions, we used two datasets: patents and economic sector data, following other studies (Pinheiro et al. 2022). It is well-known that patents cover only a small part of an economy as patenting activity varies widely even among knowledge-intensive industries. This is especially the case in a developing country like Brazil. Therefore, we also used sectoral data in order to cover the whole economy and to include sectors that would not have been represented in an analysis strictly based on patents.

Economic sectoral data were put together from the Annual Social Security Information Report (RAIS) compiled by the Labour Secretary in the Ministry of Economy. We gathered data on 87 two-digit NACE sectors (Statistical Classification of Economic Activities in the European Community) and on the number of companies in each sector in 137 functional meso-regions of Brazil following the spatial classification of IBGE (Brazilian Institute of Geography and Statistics) during the period 2006-2019. Meso-regions are defined according to the areas of influence of urban centers that concentrate certain facilities and services, attracting people from the surrounding areas.

Regarding data on patents, we resorted to Orbit Intelligence, a database that comprises the activity of different national intellectual property offices through a wide range of search engines. We used inventors' nationality as the query engine and refined our results according to the application priority date, from 2006 to 2019. The result of our search encompassed patents applied to all intellectual property offices in that period. As the data are grouped into families, there is no risk of double counting. To overcome the lack of a regional database for Brazilian patent data, we assigned patents to IBGE meso-regions on the basis of the location of the inventor. Schmoch's (2008) classification was used to classify the technological domains of patents.

It is important to mention here that our patent data is limited because the Brazilian intellectual property office does not openly provide location data of inventors. Accordingly, the 14,920 patents being analyzed are those that have also been filed at foreign offices. This number corresponds to 18% of tracks which included either a Brazilian assignee or a Brazilian inventor in the database. Although this reduces the number of observations, according to Schaefer and Liefner (2017), filing patents in offices other than the national one usually has to do with the expectations of gain from the new technology and also with organizational strategic aspects. In other words, our sample is composed of patents that are expected to generate higher economic returns and have higher strategic importance.

To collect data on population density and GDP *per capita* at the meso-region level, we used IPEA (The Institute for Applied Economic Research) data. This database gathers data from different Brazilian sources and organizes them at different spatial levels (IPEA, 2021).

3.2.Measuring relatedness and complexity

The next step was to calculate relatedness and complexity. To control for any variations, we followed Balland et al. (2019) and divided the data into three non-overlapping periods: 2006-2010, 2011-2015 and 2016-2019. The EconGeo R package (Balland, 2017) was used to calculate both relatedness and complexity for each of the three time periods in each of our 137 regions.

We follow the literature on diversification to determine the relatedness between economic sectors (Hidalgo et al. 2007), using co-location patterns from a 'region x sector' incidence matrix (see e.g. Cortinovis et al. 2017). This way, two sectors would be considered related if regions simultaneously present relative comparative advantage (RCA) in both sectors. The measure based on the co-location of RCA derives from the idea that related sectors require a similar set of capabilities. If they usually co-occur in regions, it might indicate that the set of skills needed for one of them also serves the other.

A region has RCA in a sector if this sector is overrepresented in the region when compared to the reference, Brazil. Being overrepresented means that the ratio of company share of an economic sector i in the region r is greater than that share across Brazil. This variable is subsequently turned into a binary, assuming the value of 1 when the share of an economic sector in the region is higher than in Brazil, and 0 otherwise. We assume that if RCA=1 or higher, the region has a specialization in this specific sector.

$$RCA_{r,i} = \frac{firms_{r,i}/\sum_i firms_{r,i}}{\sum_r firms_{r,i}/\sum_i \sum_r firms_{r,i}}$$

Next, we compute relatedness as a standardized measure of the frequency with which economic sectors are overrepresented in the same region. Thus, we calculated the relatedness density for each region in each sector, following Balland et al. (2019).

$$Relatedness\ density_{i,r,t} = \frac{\sum_{j \in r, j \neq i} \phi_{ij}}{\sum_{j \neq i} \phi_{ij}} * 100$$

Relatedness density in sector i in region r at time t is calculated using the relatedness $\phi_{i,j}$ of sector i to all other sectors j in which the region has RCA, divided by the total sum of sectoral relatedness of the sector i to all the other sectors j in the reference region, Brazil, at time t. This measure varies between 0 and 100. It has a maximum value of 100 when a region is specialized in all sectors to which a potential new sector (in which the region is not yet specialized) is related. With this measure, it is possible to identify in which sectors a region has a high potential to diversify. The higher the relatedness density, the higher the probability that this sector will enter the region as a new specialization, because there are many relevant capabilities present in the region.

Following the same logic, a region is considered to have relative technological advantage (RTA) in a technology if this technology is overrepresented in the region relative to the reference, Brazil. To calculate technology relatedness we adopted a slightly different strategy. Following Balland et al. (2019), we calculate relatedness based on the co-occurrence of technology domains on patent documents, using a 'technology x patent document' incidence matrix. Therefore, technology relatedness is a measure of the normalized frequency with which two technology domains appear on the same patent document. We use Schmoch (2008) classification of technology domains, which displays 35 technology categories. To calculate relatedness density, we followed the same approach as described before but using a 'region x technology' matrix.

Following other studies (e.g. Pinheiro et al. 2022), we excluded regions with less than 20 patents to prevent high RTA's based on very small absolute numbers of patents.

Finally, we calculated the complexity of each sector and technology for each of the periods. To calculate it, we follow other studies and apply the conventional method of reflections, implemented in the EconGeo R package. This measure was developed by Hidalgo and Hausmann (2009) for products and countries and was adapted by Balland and Rigby (2017) for regions and technologies. The method considers not only the diversity of activities present in a region, but also the ubiquity of activities based on how many other regions produce these activities in a competitive manner. To construct the index, the 'sector/technology x region' incidence matrix is row standardized. We subsequently multiplied the matrix by its transpose thus obtaining a new squared matrix (M). We then used the standardized second eigenvector $\vec{Q}$ of matrix M, as follows:

$$Complexity\ index_i = \frac{\vec{Q} - \langle \vec{Q} \rangle}{stdev\ (Q)} * 100$$

To control for variations between periods, we used an average of the complexity measures for sectors and technologies. We also calculated the average complexity of regions both for technologies and sectors.

3.3.Econometric approach

Following many works on regional diversification, we use a linear probability model (LPM), in which the dependent variable is 1 if the region r enters the sector/technology i in the period, and 0 otherwise. Entering means acquiring RCA/RTA (>1) in a sector/technology in which the region was not specialized in the previous period.

Our main variables of interest are relatedness density and complexity. Complexity is sector- and technology-specific. A positive complexity coefficient implies that the higher the complexity of a sector/technology the higher the probability that this sector/technology will enter the region. Moreover, we control for some region-specific features including: population density to account for urbanization economies; GDP *per capita*; variety of sectors/technologies in the region which is proxied by the number of sectors/technologies in the region represented by at least one firm or one patent; and the size of sector/technology, that is, in how many regions there is at least one firm or patent from the sector/technology (Balland et al., 2019). All independent variables are lagged by one period. We include region and time fixed-effects and, as errors are correlated, we clustered errors at the region and sector/technology levels.

We have run four different estimations. First, we estimated the full sample for both technologies and sectors. Second, we estimated a model accounting only for the 50%

most complex sectors and technologies in order to check if effects of relatedness density and complexity are different when we look only at the most complex activities. Finally, we split our sample in two (the 50% most and the 50% least complex regions) in order to see how the diversification process occurs in regions with different complexity levels.

We have the same number of observations for sectors in all periods (137 regions and 87 sectors). For technologies, this is different. In all periods, the same 35 technologies appear, but we include 30 regions for period 2 and 38 regions for period 3. This variation occurs due to the exclusion of regions with less than 20 patents, as mentioned earlier.

4. Results and discussion

4.1. Econometric modeling

We analyze the probability that a region develops a new specialization, that is, RCA/RTA, in a sector and technology, the probability that a region develops a new specialization in a high-complex sector and technology, and the probability that high and low complex regions develop new sectoral and technological specializations. We first discuss the results for sectors and then turn to the results on technologies.

Table 1 shows the results for estimations of the full sample and the 50% most complex sectors. As expected, relatedness density has a positive and statistically significant effect on the probability of a region to develop a new sectoral specialization in both models: the greater the relatedness density of a region in a given sector, the higher the probability of this region to develop a specialization in this sector in the next period. The magnitude of relatedness density is slightly smaller in the model that only includes the 50% most complex sectors but the results still suggest that, in order to develop RCA in more complex sectors, regions build on related capabilities. Complexity represents a negative and statistically significant effect in both models, suggesting that the probability of a region to develop a new specialization is smaller the greater the sector's complexity. This is as expected, as complex capabilities are more difficult to develop (Balland et al., 2019).

Table 1 – Entry model for sectors (2006–2019)

	Full sample		**50% most complex sectors**	
Constant	0.0551949*** (0.145533)	0.1737284 (0.1017726)	0.0628322** (0.0180919)	0.1092834 (0.1368655)
Relatedness density (t-1)	0.0022885*** (0.0002064)	0.0036257*** (0.0003148)	0.00213*** (0.0002575)	0.003147*** (0.0005369)

Complexity (t-1)	-0.0004799*** (0.0000904)	-0.0002346* (0.0000979)	-0.0012179*** (0.0001541)	-0.0011306*** (0.0001599)
Population density (t-1)	-0.0000248** (9.23e-06)	-0.000837 (0.0004458)	2.07e-06 (0.0000175)	0.0002684 (0.0008359)
GDP per capita (t-1)	-1.30e-07 (2.58e-07)	-2.43e-06 (1.49e-06)	8.45e-07* (3.51e-07)	-8.63e-09 (1.54e-06)
Size of sector (t-1)	-0.0000297 (0.000079)	-8.03e-06 (0.0000792)	0.0001005 (0.000114)	0.0001308 (0.0001142)
Variety of sectors in region (t-1)	-0.0002102 (0.000156)	-0.0010523 (0.0012002)	0.0000208 (0.0001526)	-0.000885 (0.0012364)
Region fixed-effects	No	Yes	No	Yes
Time fixed-effects	No	Yes	No	Yes
R^2	0.0214	0.0335	0.0300	0.0548
F Stat	51.04*** (6, 8871)	3.35*** (143, 8871)	32.7*** (6, 4901)	2.94*** (143, 4901)
Observations	16652	16652	9387	9387

#Heteroskedasticity-robust standard errors (clustered at region and sector level) are shown in parentheses.
#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001

Next, we divide the sample into two groups: the 50% most and 50% least complex regions, based on the sectors in which they are already specialized. This analysis follows Balland and Rigby (2017) and Hidalgo and Hausmann (2009), who argue that the local economic structure reflects the set of capabilities available. Regions with a higher average complexity are the ones that are expected to manage to perform more complex activities. Table 2 presents the main findings.

Table 2 – Entry model for sectors in the most and least complex regions (2006–2019)

	50% most complex regions		**50% least complex regions**	
Constant	0.0593003 (0.0305667)	0.3123494 (0.2977713)	0.0815606 *** (0.018886)	-0.4314623 (0.5644809)
Relatedness density (t-1)	0.0022711*** (0.0002711)	0.0037254*** (0.0004206)	0.0013709 ** (0.0004014)	0.0017142** (0.0005344)
Complexity (t-1)	0.0001408 (0.000118)	0.0002509* (0.000125)	-0.0014001*** (0.0001422)	-0.0012947*** (0.0001686)
Population density (t-1)	-0.0000158 (9.53e-06)	-0.0010225* (0.0004722)	6.72e-06 (0.0000995)	0.0041753 (0.0049631)

GDP per capita (t-1)	6.16e-08 (3.30e-07)	1.44e-06 (3.26e-06)	-1.29e-07 (4.20e-07)	-3.32e-06* (1.67e-06)
Size of sector (t-1)	-0.0002547* (0.0001278)	-0.0002202 (0.0001288)	0.0002187* (0.0000998)	0.0002343* (0.0001)
Variety of sectors in region) (t-1)	-0.0003988 (0.0003398)	-0.0032144 (0.0038111)	-0.0000287 (0.0001916)	0.0007959 (0.0013509)
Region fixed-effects	No	Yes	No	Yes
Time fixed-effects	No	Yes	No	Yes
R^2	0.0131	0.0263	0.0368	0.0458
F Stat	13.10*** (6, 3918)	2.27*** (73, 3918	42.14*** (6, 4888)	4.12*** (75, 4888)
Observations	7262	7262	9270	9270

#Heteroskedasticity-robust standard errors (clustered at the region and sector level) are shown in parentheses.
#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001.

Table 2 shows that both types of regions are positively impacted by relatedness density. However, the findings suggest that the most and the least complex regions are subject to different patterns of diversification in terms of complexity. The likelihood of the most complex regions to develop new sectoral specializations is positively impacted by sector complexity, at least in the fixed-effect estimation. This is not true for the least complex regions, as complexity shows a negative and statistically significant effect. This result indicates that the most complex regions in Brazil have more diverse and exclusive capabilities, which enhances their ability to develop new complex activities with greater potential economic benefits. Less complex regions are more likely to enter less complex sectors.

Table 3 presents the findings for technologies. Again, relatedness density shows a positive and statistically significant effect, with the exception of the fixed-effects estimation for the 50% most complex technologies. For the 50% most complex technologies, the magnitude of relatedness density is slightly smaller. In terms of complexity, we see that all estimations present a negative sign, but the coefficients are not statistically significant, in contrast to our analyses on sectors. Complexity of a technology does not have an effect on the probability of regions to develop new technological specializations in a region.

Table 3 – Entry models for technologies (2006–2019)

	Full sample		**50% most complex technologies**	
Constant	-0.0284689 (0.0489369)	-2.327555* (1.002673)	0.0164175 (0.0713526)	0.0405686 (1.297333)
Relatedness density (t-1)	0.0023933*** (0.0005705)	0.0019908 ** (.0006076)	0.0019846** (0.0007038)	0.0010053 (0.0008013)
Complexity (t-1)	-0.0001115 (0.0004348)	-0.0000454 (0.0004539)	-0.0008191 (0.0008016)	-0.0008384 (0.0008136)
Population density (t-1)	7.56e-07 (0.000025)	0.0007179 (0.0011734)	0.000045 (0.0000401)	-0.000993 (0.0018923)
GDP per capita (t-1)	5.17e-07 (8.10e-07)	0.0000154 (0.000012)	-1.42e-07 (1.05e-06)	0.0000164 (0.0000162)
Size of technology (t-1)	0.0043992*** (0.000935)	0.0044344*** (0.0009342)	0.0039042** (0.0011683)	0.0042211*** (0.0011711)
Stock of regional knowledge (t-1)	0.0011748 (0.0011029)	0.1838097* (0.0914242)	0.0028891 (0.0014765)	-0.0101782 (0.1156382)
Region fixed-effects	No	Yes	No	Yes
Time fixed-effects	No	Yes	No	Yes
R^2	0.0243	0.0812	0.0321	0.1072
F Stat	9.75*** (6, 1506)	9.37*** (59, 1506)	6.37*** (6, 800)	4.75*** (59, 800)
Observations	2204	2204	1180	1180

#Heteroskedasticity-robust standard errors (clustered at the region and sector level) are shown in parentheses.
#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001

Table 4 presents the findings when splitting the sample into the most and least complex regions according to their technological endowment. We observe different diversification patterns between the two groups. The relatedness density coefficient is positive in both cases, but it is only statistically significant for the least complex regions: relatedness is important for technological diversification in low-complexity regions but not in high-complexity regions. The complexity variable is positive and significant for technological diversification in the most complex regions but negative and significant for the least complex regions. That is to say, the higher the complexity of a technology the lower the probability that this technology will enter the low-complexity region, in contrast to the case of high-complexity regions. We found similar findings for sectoral diversification reported earlier in this section.

Table 4 - Entry model for technologies in the most and least complex regions (2006-2019)

	50% most complex regions		**50% least complex regions**	
Constant	-0.0132546 (0.0895805)	-0.1051717 (0.1765559)	0.1982948* (0.0909253)	0.2073085 (0.5865322)
Relatedness density (t-1)	0.0003471 (0.0010003)	-0.0001913 (0.0010487)	0.0047638*** (0.0009236)	0 .0035174** (0.0010197)
Complexity (t-1)	0.0031818*** (0.0007999)	0.0033978*** (0.0008229)	-0.0030147*** (0.0007061)	-0.0031472*** (0.0007241)
Population density (t-1)	-0.0000474 (0.0000384)	-0.0001967 (0.0017637)	6.68e-06 (0.0000344)	0.0001112 (0.0016023)
GDP per capita (t-1)	-1.52e-06 (1.67e-06)	0.0000142 (0.000016)	-2.92e-08 (1.04e-06)	6.11e-07 (0.0000176)
Size of technology (t-1)	0.0040832* (0.0016775)	0.0042716* (0.0017137)	0.0025112 (0.0015611)	0.0025843 (0.0015679)
Stock of regional knowledge (t-1)	0.0001078 (0.0018019)	-0.0050766 (0.010956)	-0.0006323 (0.0019855)	0.0023968 (0.0212229)
Region fixed-effects	No	Yes	No	Yes
Time fixed-effects	No	Yes	No	Yes
R^2	0.0254	0.0601	0.0823	0.1049
F Stat	3.53*** (6, 501)	2.24*** (23, 501)	12.07*** (6, 499)	4.94*** (22, 499)
Observations	810	810	816	816

#Heteroskedasticity-robust standard errors (clustered at the region and economic sector level) are shown in parentheses.
#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001

In sum, results point out that different types of regions are characterized by different patterns of diversification. Relatedness density turns out to be important for technological diversification in low-complexity regions but not in high-complexity regions. This is different for sectoral diversification where relatedness density mattered also in high-complex regions. Complexity shows a consistent pattern: it is negatively correlated with technological and sectoral diversification in less complex regions but positively correlated in more complex regions.

4.2. Diversification opportunities of regions

To further assess how diversification opportunities differ between types of regions, we classified the 137 Brazilian meso-regions into three groups: central, intermediate and peripheral regions. The first group is composed of 24 regions, mainly metropolitan areas of state capitals; the second is composed of 18 regions, mainly from the southeastern and southern states; and the last group is composed of 95 regions from all over the country. To make this classification, we adapted Marsan and Maguire (2011),[1] using the variables available at the regional level for Brazil. We used the average data from the period 2006 to 2019 for the following variables: agriculture, industry and services GDP, the share of the population with tertiary education, population density, GDP *per capita*, applied patents per million inhabitants (based on our patents database) and average regional complexity. Next, we conducted a cluster analysis, based on the Canberra distance in which clusters are formed according to the distance between pairs of points in a vector space (Faisal et al., 2020).

We make use of the framework of Balland et al. (2019) to identify diversification opportunities of each region in terms of relatedness and complexity. For illustrative purposes, we took a representative example of each of the three types of regions: São Paulo Metropolitan Area as an example of a central region; Northeast of Rio Grande do Sul as an example of an intermediate region; and São Francisco Pernambucano as an example of a peripheral region. The diversification opportunities of each of the three regions are plotted for both sectors and technologies in Figures 1, 2 and 3[2]. Each dot represents a sector or a technology in which the region has RCA/RTA or not. The green dots represent the sectors and technologies in which the region is already specialized, and the red dots are the ones in which the region is not specialized. That is to say, the red dots represent potential entrants in the regions and score high or low on relatedness and complexity.

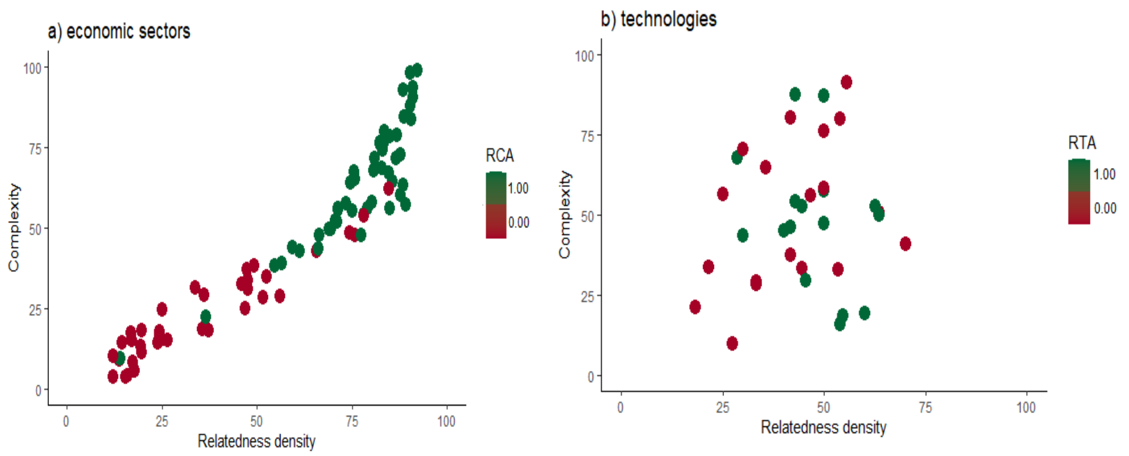
The case of São Paulo Metropolitan Area is presented in Figure 1. A clear pattern emerges for sectors. The sectors in which the region is already specialized (RCA=1) show a high level of relatedness density. As expected, relevant and supportive capabilities

[1] The authors employed cluster techniques to classify the OECD regions in 8 groups, according to 12 socio-economic and innovation-related variables.

[2] It is worth mentioning that only two regions from the peripheral group had patent applications in our database.

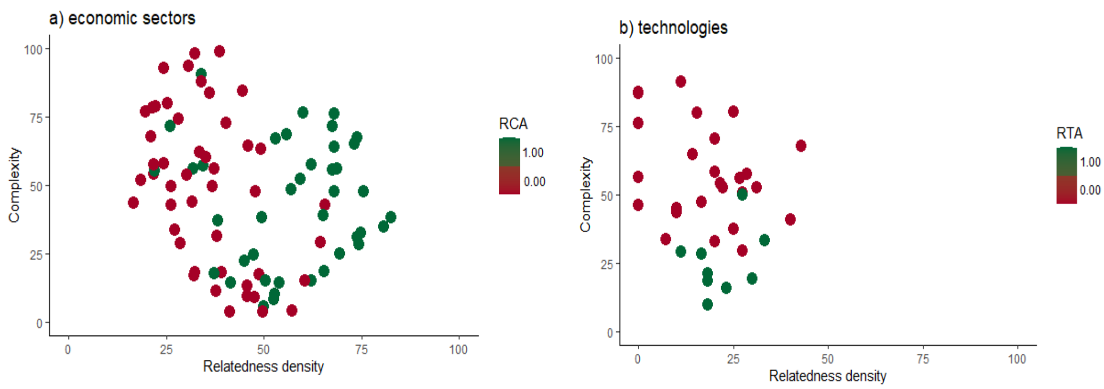
(related sectors) have a high presence in the region. Luckily for the region, these also turn out to be the sectors with the highest complexity. Moreover, there is a positive relationship between relatedness and complexity, implying that the region has most diversification opportunities (RCA=0) in complex sectors (there are a few red dots with high relatedness density such as manufacture of other transport equipment and health activities), but least opportunities in non-complex sectors, as reflected in their low relatedness density. For technologies, the relationship between relatedness and complexity is less straightforward. Existing specializations (RTA=1) are both in low and high-complex technologies. The same applies for diversification opportunities (RTA=0). Figure 1b shows there are many opportunities for diversifying into complex technologies, in which the São Paulo Metropolitan Area has high relatedness density, such as computer technologies and semiconductors. Policy could focus on those potential sectors and technologies (RCA/RTA=0) in the region that score high on both complexity and relatedness density. The São Paulo Metropolitan area may conduct a related diversification policy approach (Balland et al. 2019), targeting complex activities in which they have strong local capabilities.

Figure 1 – Diversification opportunities in São Paulo Metropolitan Area (central region)



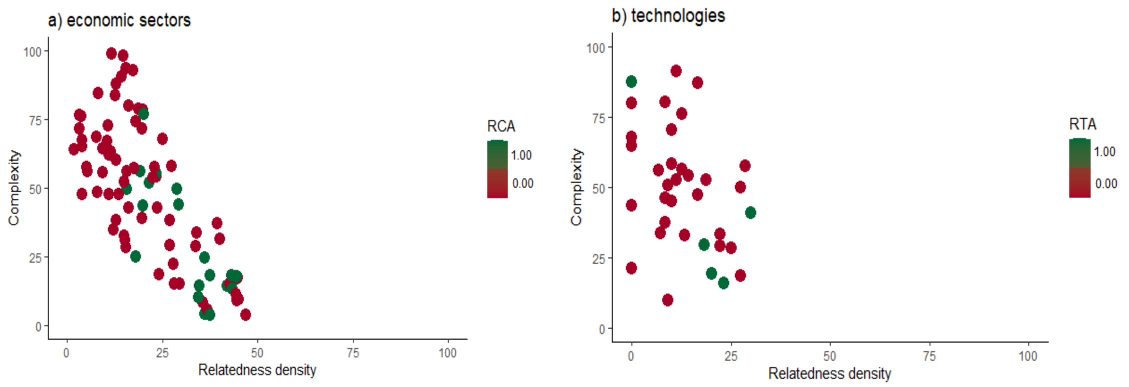
The case of the Northeast of Rio Grande do Sul is presented as an example of an intermediate region in Figure 2. The region appears to be specialized in both low and high complex sectors (RCA=1). There is no clear positive relationship between relatedness and complexity as in the São Paulo Metropolitan Area. As expected, most of those sectors score relatively higher on relatedness as compared to sectors in which the region is not specialized (RCA=0). There are still some diversification opportunities in complex sectors in the region, such as manufacture of pharmaceutical products. Regarding technologies, the region is specialized in only a few technologies (RTA=1), almost all of them low complexity technologies. Prospects for diversification in complex technologies are relatively low, as most of them show a relatedness density of lower than 25. All in all, this region has some potential to move into complex sectors and technologies but these opportunities are much lower than the São Paulo Metropolitan Area. More opportunities exist for low complexity activities which could be a second-best policy option.

Figure 2 – Diversification opportunities in Northeast of Rio Grande do Sul (intermediate region)



The case of São Francisco Pernambucano is presented in Figure 3. This peripheral region sharply contrasts the cases of the two other regions. For sectors, Figure 3a shows a negative relationship between relatedness and complexity. The region is primarily specialized in low-complex sectors, and diversification opportunities are in low-complex rather than high-complex sectors. Looking at technologies, the region has a limited number of specializations. There is also a shortage of diversification opportunities: for almost all technologies, relatedness density scores are low. While being low in general, diversification opportunities are even lower for high-complex technologies. Such a picture is indicative of the challenge to construct policies that aim to enhance the ability of peripheral regions to diversify into more complex activities (Boschma, 2022a).

Figure 3 – Diversification opportunities in São Francisco Pernambucano (peripheral region)



In sum, the diversification process is very region-specific. Diversification opportunities of regions turn out to be very different. We found striking differences between the three categories of regions, but also across individual regions.

5. Conclusion

This article aimed to investigate the sectoral and technological diversification process in an emerging economy context and to explore different diversification patterns that occur in regions with varying levels of complexity levels, taking Brazil as case.

We found that relatedness enhances the ability of regions to diversify into new sectors and new technologies. In other words, sectors and technologies requiring similar capabilities to those available in the regional portfolio are more likely to enter the region. This is in line with other studies (Alonso & Martín, 2019; Boschma, 2017; Galetti et al., 2021). There is one exception: relatedness is not as important for technological diversification in high-complexity regions. This comes close to what other studies on diversification found at the national (Petralia et al. 2017; Pinheiro et al. 2021) and the regional scale (Xiao et al., 2018; Zhu et al., 2017). More advanced regions are less impacted by relatedness when diversifying. This may be attributed to their strong research and innovation capacity (Xiao et al., 2018; Zhu et al., 2017) and the role of foreign investments (Breul & Pruss, 2022; He et al., 2018).

Broadly speaking, we found that complexity is negatively correlated with regional diversification: the higher the complexity of a sector or a technology, the lower the probability that it will enter a region. This reflects the fact that complex activities are more difficult to develop (Balland et al. 2019). In high-complexity regions, however, the opposite is true: complexity now turns into a positive effect for both sectoral and technological diversification. These results seem to be in line with recent studies at the national (Hartmann et al. 2021; Pinheiro et al. 2021) and regional scale (Pinheiro et al. 2022) that argued that low complexity economies are more likely to enter low complexity products, as they are more related to them, while more complex economies are more likely to diversify towards complex products.

Our study underlines that regions with varying complexity levels are characterized by different types of diversification. This affects not only the role of relatedness in the diversification process, but also impacts the tendency of regions to diversify in low or high-complex activities. Our study also demonstrated that the diversification opportunities of regions look very different. We reported striking differences between three categories of regions (central, intermediate, peripheral). Central regions – corresponding to more developed urban areas – have the best opportunities to diversify in complex sectors. Peripheral (less developed) regions have diversification opportunities only in low-complexity sectors. Central regions also show more opportunities to develop new complex technologies than intermediate regions where most opportunities are in low-complex activities. The capabilities of peripheral regions leave them with little opportunities to develop new technologies.

These findings have policy implications. Much debate on regional diversification opportunities has centered around Smart Specialization policies in Europe (Balland et al., 2019; Foray, 2016; McCann & Ortega-Argilés, 2015). Hartmann et al. (2021) argued that emerging economies that adopted smart policies in the past, such as South Korea and Singapore, were the ones that managed to move up to more complex activities. The justification for adopting such policies in emerging economies, according to Hartmann et al. (2021), is the fact that these countries are more related to low-complex products, however, at the same time, they have the basic skills to produce more complex and high-value-added activities. Our study shows such policy should be made region-specific, as local capabilities provide opportunities but also set limits to what can be achieved by Smart Specialization policies to diversify into low or complex activities.

Needless to say, our findings have limitations and also open up new research questions. First, although we differentiate regions according to their degree of complexity, it would be interesting to explore which other characteristics determine the type of prevailing diversification pattern. Moreover, our analysis was focused on the local endowments of regions. We did not explore how regions could benefit from external inputs, such as collaboration with other regions (Balland & Boschma, 2021; Barzotto et al., 2019), FDI (Alonso & Martín, 2019) or global value chains (Boschma, 2022b). This investigation would be relevant especially for peripheral regions, which suffer from a limited stock of local capabilities (Balland & Boschma, 2021).

References

Abramovitz, M. (1986). Catching Up, Forging Ahead, and Falling Behind. *The Journal of Economic History*, *46*(2). https://doi.org/10.1017/S0022050700046209

Aghion, P., & Howitt, P. (1992). A Model of Growth Through Creative Destruction. *Econometrica*, *60*(2). https://doi.org/10.2307/2951599

Alonso, J. A., & Martín, V. (2019). Product relatedness and economic diversification at the regional level in two emerging economies: Mexico and Brazil. *Regional Studies*, *53*(12). https://doi.org/10.1080/00343404.2019.1605441

Andersson, M., Bjerke, L., & Karlsson, C. (2013). Import flows: Extraregional linkages stimulating renewal of regional sectors? *Environment and Planning A*, *45*(12). https://doi.org/10.1068/a45732

Antonelli, C., Crespi, F., & Quatraro, F. (2020). Knowledge complexity and the mechanisms of knowledge generation and exploitation: The European evidence. *Research Policy*. https://doi.org/10.1016/j.respol.2020.104081

Audretsch, D. B., & Feldman, M. P. (1996). R&D Spillovers and the Geography of Innovation and Production. *American Economic Review*, *86*(3). https://doi.org/10.2307/2118216

Balland, P. A. (2017). *Economic Geography in R: Introduction to the EconGeo package* (No. 17; 09).

Balland, P.A., & Boschma, R. (2021). Complementary interregional linkages and Smart Specialisation: an empirical study on European regions. *Regional Studies*, *55*(6). https://doi.org/10.1080/00343404.2020.1861240

Balland, P.A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, *53*(9). https://doi.org/10.1080/00343404.2018.1437900

Balland, P.A., Jara-Figueroa, C., Petralia, S. G., Steijn, M. P. A., Rigby, D. L., & Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour*, *4*(3). https://doi.org/10.1038/s41562-019-0803-3

Balland, P.A., & Rigby, D. (2017). The Geography of Complex Knowledge. *Economic Geography*, *93*(1). https://doi.org/10.1080/00130095.2016.1205947

Barzotto, M., Corradini, C., Fai, F. M., Labory, S., & Tomlinson, P. R. (2019). Enhancing innovative capabilities in lagging regions: An extra-regional collaborative approach to RIS3. *Cambridge Journal of Regions, Economy and Society*, *12*(2). https://doi.org/10.1093/cjres/rsz003

Boschma, R. (2017). Relatedness as driver of regional diversification: a research agenda. *Regional Studies*, *51*(3). https://doi.org/10.1080/00343404.2016.1254767

Boschma, R. (2022a). Designing Smart Specialization Policy: relatedness, unrelatedness, or what? In M. Anderson, C. Karlsson, & S. Wixe (Eds.), *Handbook of Spatial Diversity and Business Economics*. Oxford University Press.

Boschma, R. (2022b). Global Value Chains from an Evolutionary Economic Geography perspective: a research agenda. *Area Development and Policy*. https://doi.org/10.1080/23792949.2022.2040371

Boschma, R. A. (2004). Competitiveness of regions from an evolutionary perspective. *Regional Studies*, *38*(9). https://doi.org/10.1080/0034340042000292601

Boschma, R., Balland, P. A., & Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, *24*(1). https://doi.org/10.1093/icc/dtu012

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research Policy*, *32*(1). https://doi.org/10.1016/S0048-7333(02)00004-5

Breul, M., & Pruss, F. (2022). Applying Evolutionary Economic Geography beyond case studies in the Global North: Regional diversification in Vietnam. *Singapore Journal of Tropical Geography*, 43(1), 26-42.

Colombelli, A., Krafft, J., & Quatraro, F. (2014). The emergence of new technology-based sectors in European regions: A proximity-based analysis of nanotechnology. *Research Policy*, *43*(10). https://doi.org/10.1016/j.respol.2014.07.008

Cortinovis, N., Xiao, J., Boschma, R., & van Oort, F. G. (2017). Quality of government and social capital as drivers of regional diversification in Europe. *Journal of Economic Geography*, *17*(6). https://doi.org/10.1093/jeg/lbx001

Davies, B., & Maré, D. C. (2021). Relatedness, complexity and local growth. *Regional Studies*, *55*(3). https://doi.org/10.1080/00343404.2020.1802418

Dosi, G. (1982). Technological paradigms and technological trajectories. A suggested interpretation of the determinants and directions of technical change. *Research Policy*, *11*(3). https://doi.org/10.1016/0048-7333(82)90016-6

Dosi, G. (1988). The nature of the innovative process. In *Technical Change and Economic Theory*.

Essletzbichler, J. (2015). Relatedness, Industrial Branching and Technological Cohesion in US Metropolitan Areas. *Regional Studies*, *49*(5). https://doi.org/10.1080/00343404.2013.806793

Fagerberg, J., & Srholec, M. (2018). National innovation systems, capabilities and economic development. In *Innovation, Economic Development and Policy: Selected Essays*. https://doi.org/10.1016/j.respol.2008.06.003

Faisal, M., Zamzami, E. M., & Sutarman. (2020). Comparative Analysis of Inter-Centroid K-Means Performance using Euclidean Distance, Canberra Distance and Manhattan Distance. *Journal of Physics: Conference Series*, *1566*(1). https://doi.org/10.1088/1742-6596/1566/1/012112

Feldman, M. P., & Kogler, D. F. (2010). Stylized facts in the geography of innovation. In *Handbook of the Economics of Innovation* (Vol. 1, Issue 1 C). https://doi.org/10.1016/S0169-7218(10)01008-7

Foray, D. (2016). On the policy space of smart specialization strategies. *European Planning Studies*, *24*(8). https://doi.org/10.1080/09654313.2016.1176126

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, *41*(5). https://doi.org/10.1080/00343400601120296

Galetti, J. R. B., Tessarin, M. S., & Morceiro, P. C. (2021). Skill relatedness, structural change and heterogeneous regions: evidence from a developing country. *Papers in Regional Science*. https://doi.org/10.1111/pirs.12629

Greene, W. W. H. . (2012). Econometric analysis 7th Ed. In *Prentice Hall* (Vol. 97).

Hartmann, D., Zagato, L., Gala, P., & Pinheiro, F. L. (2021). Why did some countries catch-up, while others got stuck in the middle? Stages of productive sophistication and smart industrial policies. *Structural Change and Economic Dynamics*, *58*. https://doi.org/10.1016/j.strueco.2021.04.007

Hausmann, R., Hidalgo, C. A., Bustos, S. A., Coscia, M., Simoes, A., & Yildirim, M.

A. (2014). The Atlas of economic complexity: mapping paths to prosperity. *Choice Reviews Online*, *51*(11). https://doi.org/10.5860/choice.51-5931

He, C., Yan, Y., & Rigby, D. (2018). Regional industrial evolution in China. *Papers in Regional Science*, *97*(2). https://doi.org/10.1111/pirs.12246

Hidalgo, C. A., Winger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, *317*(5837). https://doi.org/10.1126/science.1144581

Hidalgo, César A., Balland, P. A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., & Zhu, S. (2018). The Principle of Relatedness. In *Springer Proceedings in Complexity*. https://doi.org/10.1007/978-3-319-96661-8_46

Hidalgo, César A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26). https://doi.org/10.1073/pnas.0900943106

IPEA. (2021). *IPEA Data*. Instituto de Pesquisa Econômica Aplicada. http://www.ipeadata.gov.br/

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, *108*(3). https://doi.org/10.2307/2118401

Jara-Figueroa, C., Jun, B., Glaeser, E. L., & Hidalgo, C. A. (2018). The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(50). https://doi.org/10.1073/pnas.1800475115

Keun, L. (2012). Schumpeterian analysis of economic catch-up: Knowledge, path-creation, and the middle-income trap. In *Schumpeterian Analysis of Economic Catch-Up: Knowledge, Path-Creation, and the Middle-Income Trap*. https://doi.org/10.1017/CBO9781107337244

Lall, S. (1992). Technological capabilities and industrialization. *World Development*, *20*(2). https://doi.org/10.1016/0305-750X(92)90097-F

Marsan, G. A., & Maguire, K. (2011). Categorisation of OECD Regions Using Innovation-Related Variables. In *OECD Regional Development Working Papers*.

McCann, P., & Ortega-Argilés, R. (2015). Smart Specialization, Regional Growth and Applications to European Union Cohesion Policy. *Regional Studies*, *49*(8). https://doi.org/10.1080/00343404.2013.799769

Mewes, L., & Broekel, T. (2020). Technological complexity and economic growth of regions. *Research Policy*. https://doi.org/10.1016/j.respol.2020.104156

Neffke, F., Hartog, M., Boschma, R., & Henning, M. (2018). Agents of Structural Change: The Role of Firms and Entrepreneurs in Regional Diversification. *Economic Geography*, *94*(1). https://doi.org/10.1080/00130095.2017.1391691

Neffke, F., Henning, M., & Boschma, R. (2011). How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography*, *87*(3). https://doi.org/10.1111/j.1944-8287.2011.01121.x

Nelson, R. R., & Winter, S. G. (1982). An Evolutionary Theory of Change. In *Massachusetts, Belknap Harvard*.

Nooteboom, B. (2011). Learning and Innovation in Organizations and Economies. In *Learning and Innovation in Organizations and Economies*. https://doi.org/10.1093/acprof:oso/9780199241002.001.0001

Petralia, S., Balland, P. A., & Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy*, *46*(5). https://doi.org/10.1016/j.respol.2017.03.012

Pinheiro, F. L., Balland, P. A., Boschma, R., & Hartmann, D. (2022). *The Dark Side of the Geography of Innovation: Relatedness, Complexity, and Regional Inequality in Europe* (22.02; Papers in Evolutionary Economic Geography). http://econ.geo.uu.nl/peeg/peeg2202.pdf

Pinheiro, F. L., Hartmann, D., Boschma, R., & Hidalgo, C. A. (2021). The time and frequency of unrelated diversification. *Research Policy*. https://doi.org/10.1016/j.respol.2021.104323

Pintar, N., & Scherngell, T. (2020). The complex nature of regional knowledge production: Evidence on European regions. *Research Policy*. https://doi.org/10.1016/j.respol.2020.104170

Rigby, D. L. (2015). Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes. *Regional Studies*, *49*(11). https://doi.org/10.1080/00343404.2013.854878

Rigby, D., Roesler, C., Kogler, D. F., Boschma, R., & Balland, P. A. (2022). Do EU regions benefit from Smart Specialization principles? *Regional Studies*. https://doi.org/10.1080/00343404.2022.2032628

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, *98*(5). https://doi.org/10.3386/w3210

Romero, J. P., Freitas, E., Silveira, G., Britto, F., Cimini, F., & Jayme, G. (2021). *Economic complexity and regional economic development: evidence from Brazil*.

Schaefer, K. J., & Liefner, I. (2017). Offshore versus domestic: Can EM MNCs reach higher R&D quality abroad? *Scientometrics*, *113*(3). https://doi.org/10.1007/s11192-017-2533-5

Schmoch, U. (2008). Concept of a technology classification for Country comparison. *Final Report to the World Intellectual Property Organization*, *June*.

Teece, D. J. (1992). Competition, cooperation, and innovation. Organizational arrangements for regimes of rapid technological progress. *Journal of Economic Behavior and Organization*, *18*(1). https://doi.org/10.1016/0167-2681(92)90050-L

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. *Booksgooglecom*, *58*(2). https://doi.org/10.1515/humr.2003.021

Xiao, J., Boschma, R., & Andersson, M. (2018). Industrial Diversification in Europe: The Differentiated Role of Relatedness. *Economic Geography*, *94*(5). https://doi.org/10.1080/00130095.2018.1444989

Yeung, H. W. chung. (2021). Regional worlds: from related variety in regional diversification to strategic coupling in global production networks. *Regional Studies*, *55*(6). https://doi.org/10.1080/00343404.2020.1857719

Zhu, S., Hey, C., & Zhou, Y. (2017). How to jump further and catch up? Path-breaking in an uneven industry space. *Journal of Economic Geography*, *17*(3). https://doi.org/10.1093/jeg/lbw047

Appendix

In this section, we conduct some robustness checks to verify if our results hold if we adopt different estimation strategies.

A. Robustness check: Changing the splitting criterium of most and least complex regions

In Section 4.1, we split our sample into two groups, the 50% most and the 50% least complex regions, to show how different regions are affected by our variables of interest, that is, relatedness density and complexity. Here, we change our split criterium taking the 25% most and the 25% least complex regions. Results are shown in Tables A1 and A2. The tables show our results hold, as signs and statistical significance are similar to the results presented in Section 4.1.

Table A1 – Entry model for sectors in most and least complex regions (2006-2019)

	25% most complex regions		25% least complex regions	
Constant	-0.2872806 (0.169374)	-0.360451 (0.6565116)	0.0702191** (0.0251518)	0.1460994 (0.5951053)
Relatedness density (t-1)	0.0026253*** (0.0004608)	0 .0042218*** (0.0006143)	0.0021701*** (0.0005933)	0.0032183*** (0.0008007)
Complexity (t-1)	0.0009308*** (0.0001855)	0.0008013*** (0.000191)	-0.0014084*** (0.0002077)	-0.0010437*** (0.000249)
Population density (t-1)	-0.000021* (0.0000107)	-0.0008762 (0.0005328)	0.0002212 (0.0002681)	-0.0017454 (0.0125965)
GDP per capita (t-1)	4.58e-09 (4.09e-07)	3.72e-06 (5.24e-06)	-3.47e-08 (5.90e-07)	-4.54e-06 (2.39e-06)
Size of economic sector (t-1)	-0.0005479** (0.0002078)	-0.0004704* (0.0002082)	0.0000575 (0.0001494)	0.0000639 (0.0001494)
Variety (t-1)	0.0036206 (0.0020183)	0.0040011 (0.0073066)	0.0000991 (0.0002371)	0.0003228 (0.0016166)
Region fixed-effects	No	Yes	No	Yes
Time fixed-effects	No	Yes	No	Yes
R^2	0.0268	0.0455	0.0554	0.0654
F Stat	11.85*** (6, 1705)	3.19*** (39, 1705)	25.49*** (6, 2427)	4.31*** (40, 2427)
Observations	3115	3115	4611	4611

#Heteroskedasticity-robust standard errors (clustered at the region and sector level) are shown in parentheses.
#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001

Table A2 – Entry model for technologies in most and least complex regions (2006–2019)

	25% most complex regions		25% least complex regions	
Constant	-0.1285447 (0.135259)	-0.8767025 (0.4619315)	0.3575936* (0.1394273)	0.4534052 (0.6097529)
Relatedness density (t-1)	0.0021903 (0.0014177)	0.0015405 (0.0014124)	0.0035267* (0.0014076)	0.0029568* (0.0014267)
Complexity (t-1)	0.0050515*** (0.0009999)	0.0054672*** (0.0010394)	-0.003839*** (0.0009364)	-0.0038451*** (0.0009634)
Population density (t-1)	-0.0000288 (0.0000599)	-0.0012073 (0.0019737)	-0.0006635 (0.0006589)	-0.0073986 (0.0098314)
GDP per capita (t-1)	-3.73e-06 (4.98e-06)	0.0000185 (0.000021)	-7.04e-06 (4.60e-06)	-9.70e-06 (0.000034)
Size of technology (t-1)	0.0053538* (0.0020619)	0.0058025** (0.0020927)	0.0030764 (0.0021166)	0.0030862 (0.0021496)
Stock of regional knowledge (t-1)	-0.0013671 (0.0026474)	0.0127492 (0.0074849)	0.0045061 (0.0049813)	0.0261788 (0.0264236)
Region fixed-effects	No	Yes	No	Yes
Time fixed-effects	No	Yes	No	Yes
R^2	0.0698	0.1113	0.1091	0.1235
F Stat	6.79 (6, 280)	4.23 (15, 280)	9.84*** (6, 241)	5.09*** (13, 241)
Observations	457	457	400	400

#Heteroskedasticity-robust standard errors (clustered at the region and sector level) are shown in parentheses.
#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001

B. Robustness check: Probit estimation

Our entry variable is a binary, assuming the value of 1 when a new specialization enters the region, and 0 if it does not. Cases where the dependent variable is binary are usually estimated by nonlinear models, such as logit and probit (Wooldridge, 2002). However, these models' outcomes may be biased and inconsistent when there is a large number of dummy variables (Cortinovis et al., 2017; Greene, 2012). We re-estimate all models using the probit model, as a robustness check.

Table B1 – Entry model for sectors and technologies (2006–2019)

	Full sample economic sectors	**50% most complex sectors**	**Full sample technologies**	**50% most complex technologies**
Constant	-1.596223*** (0.1162946)	-1.40547*** (0.2103643)	-169638*** (0.1967286)	-1.581611*** (0.3034142)
Relatedness density (t-1)	0.0166299*** (0.0012942)	0.0183134*** (0.0018509)	0.0082163 *** (0.0019712)	0.0074155** (0.0026828)
Complexity (t-1)	-0.004582*** (0.00068)	-0.0166558*** (0.0020091)	-0.0004012 (0.001578)	-0.0031613 (0.0031779)
Population density (t-1)	-0.0001755* (0.0000719)	-0.0000412 (0.0001044)	2.65e-06 (0.0000779)	0.0001388 (0.0001147)
GDP per capita (t-1)	9.90e-08 (1.76e-06)	6.95e-06** (2.49e-06)	1.82e-06 (2.72e-06)	-3.86e-07 (4.09e-06)
Size of sector/ technology (t-1)	-0.0002212 (0.0005578)	0.0013934 (0.0009354)	0.0163026*** (0.0035345)	0.0151446** (0.0044352)
Variety (t-1)	-0.0016695 (0.001382)	0.0019292 (0.0023325)	0.0046668 (0.0037972)	0.0117346* (0.0054678)
(Pseudo) R^2	0.0438	0.0746	0.0242	0.0336
Observations	16652	9387	2204	1180

#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001

Table B1, we can see our results for both economic sectors and technologies hold when we adopt a different estimation method, as signs and statistical significance are similar to the results presented in section 4.1.

Table B2 - Entry model for sectors and technologies in the most and least complex regions (2006 – 2019)

	50% most complex regions (sectors)	**50% least complex regions (sectors)**	**50% most complex regions (technologies)**	**50% least complex regions (technologies)**
Constant	-1.529806 *** (0.2219166)	-117495*** (0.1820735)	-1.064074*** (0.2894409)	-1.341238*** (0.3164793)
Relatedness density (t-1)	0.0152683*** (0.0016975)	0.0055783 (0.0030746)	0.0023198 (0.0031721)	0.0175697*** (0.0032625)
Complexity (t-1)	0.000657 (0 .0008722)	-0.0153023*** (0.0014817)	0.0043603* (0.0020494)	-0.0051474* (0.0021039)
Population density (t-1)	-0.0001165 (0.0000732)	0.0003403 (0.0008576)	-0.0001725 (0.0001348)	0.0000255 (0.0001002)
GDP per capita (t-1)	7.78e-07 (2.06e-06)	-3.74e-07 (4.23e-06)	-4.79e-06 (5.87e-06)	1.64e-07 (3.53e-06)

Size of sector/ technology (t-1)	-0.0015906* (0.0007831)	0.0014115 (0.0008152)	0.0058208 (0.0051822)	0.0144458** (0.005321)
Variety (t-1)	-0.0033182 (0.002687)	-0.0001479 (0.0019099)	-0.00047 (0.006004)	-0.0031047 (0.0071587)
(Pseudo) R^2	0.0233	0.0846	0.0102	0.0661
Observations	7262	9270	810	816

#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001

In Table B2 we see that, although the signs of relatedness density and complexity are the same as in the OLS estimations, the statistical significance is not the same for relatedness density in the sectoral diversification for the least complex regions, as the variable is statistically significant in the OLS estimation.

C. Robustness check: changing the dependent variable.

Finally, following He et al. (2018) and Zhu et al. (2017), we re-estimate all models using a different threshold value to determine the RCA/RTA. We adopt the value of 0.8. Results are displayed in the following tables.

Table C1 - Entry model for sectors (2006–2019)

	Full sample		**50% most complex sectors**	
Constant	-0.0303639 (0.0170992)	-0.1675058 (0.1343223)	0.0148627 (0.0230175)	-0.1291907 (0.1856395)
Relatedness density (t-1)	0.002273*** (0.0002002)	0.0032228*** (0.0002934)	0.002671*** (0.0002642)	0.0040118*** (0.0005451)
Complexity (t-1)	-0.0002764* (0.0001338)	-0.0000324 (0.0001457)	-0.0018623*** (0.0002325)	-0.0018506*** (0.0002353)
Population density (t-1)	-0.0000553*** (0.0000121)	0.000346 (0.0005331)	-0.0000356 (0.0000258)	0.0004845 (0.0010706)
GDP per capita (t-1)	-5.36e-07 (3.20e-07)	1.59e-06 (2.07e-06)	1.88e-07 (4.36e-07)	3.51e-06 (2.49e-06)
Size of sector (t-1)	0.0005316*** (0.000096)	0.0005515*** (0.0000971)	0.0009142*** (0.0001392)	0.0009937*** (0.0001422)
Variety (t-1)	0.0003041 (0.0001908)	0.0016309 (0.001549)	0.0003472 (0.0002069)	0.0019075 (0.0017739)

Region fixed-effects	No	Yes	No	Yes
Time fixed-effects	No	Yes	No	Yes
R^2	0.0235	0.0394	0.0373	0.0681
F Stat	47.44*** (6, 7338)	4.37*** (143, 7338)	39.79*** (6, 4267)	not reported
Observations	13383	13383	7933	7933

#Heteroskedasticity-robust standard errors (clustered at the region and sector level) are shown in parentheses.
#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001

Our variables of interest presented similar results in terms of sign and statistical significance. The only difference is that adopting 0.8 as the RCA threshold, complexity was not significant in the full sample fixed-effects estimation.

Table C2 – Entry model for sectors in the most and least complex regions (2006–2019)

	50% most complex regions		**50% least complex regions**	
Constant	-0.0102348 (0.0348168)	-0.2325036 (0.3359358)	-0.0078228 (0.0227298)	0.0789034 (0.8302346)
Relatedness density (t-1)	0.0021742 (0.0002755)	0.0032499*** (0.0004049)	0.0018254*** (0.0003616)	0.0018092*** (0.0004648)
Complexity (t-1)	0.0003286 (0.0001696)	0.0004546* (0.0001801)	-0.0013035*** (0.0002076)	-0.0012972*** (0.0002425)
Population density (t-1)	-0.0000433*** (0.0000124)	0.0002182 (0.0005526)	0.0003016* (0.0001495)	-0.0012645 (0.0073505)
GDP per capita (t-1)	-4.30e-07 (3.89e-07)	-2.39e-06 (4.22e-06)	4.69e-07 (6.59e-07)	2.91e-06 (2.37e-06)
Size of sector (t-1)	0.0004505** (0.0001392)	0.0005087*** (0.0001417)	0.0006909*** (0.0001304)	0.0007166*** (0.0001312)
Variety (t-1)	-0.0002332 (0.0004258)	0.0020593 (0.0042225)	0.0004543 (0.0002386)	0.0017202 (0.0019101)
Region fixed-effects	No	Yes	No	Yes
Time fixed-effects	No	Yes	No	Yes
R^2	0.0131	0.0317	0.0400	0.0485
F Stat	13.85*** (6, 3629)	4.28*** (85, 3629)	35*** (6, 3708)	4.31*** (64, 3708)
Observations	6509	6509	6874	6874

#Heteroskedasticity-robust standard errors (clustered at the region and sector level) are shown in parentheses.
#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001

Turning to the most and least complex regions, we see that our variables of interest presented similar results in terms of sign and statistical significance. The only difference is that relatedness density was not statistically significant for the most complex regions in one of the estimations. For technologies, we see that our results hold for a different RTA threshold, with similar results for relatedness density and complexity in terms of sign and statistical significance.

Table C3 - Entry model for technologies (2006–2019)

	Full sample		**50% most complex technologies**	
Constant	-0.066795 (0.0527153)	-2.669978* (1.110629)	-0.0062228 (0.0753722)	0.0929953 (1.463824)
Relatedness density (t-1)	0.0022541*** (0.0005811)	0.0021019** (0.0006126)	0.0020346** (0.0007005)	0.0009464 (0.0007824)
Complexity (t-1)	-0.0003163 (0.0004859)	-0.0001708 (0.0004982)	-0.000647 (0.0008595)	-0.0005422 (0.0008568)
Population density (t-1)	0.0000109 (0.0000305)	0.0008621 (0.0013066)	0.0001047* (0.000051)	-0.0025887 (0.0021439)
GDP per capita (t-1)	7.37e-07 (8.81e-07)	0.0000185 (0.0000134)	-9.78e-07 (1.08e-06)	0.0000148 (0.0000172)
Size of technology (t-1)	0.0053811*** (0.0010118)	0.0055553*** (0.001005)	0.0042748** (0.001257)	0.0047711*** (0.0012774)
Stock of regional knowledge (t-1)	0.00326* (0.0013757)	0.2093516* (0.1011927)	0.0041354* (0.0017106)	-0.0047201 (0.129466)
Region fixed-effects	No	Yes	No	Yes
Time fixed-effects	No	Yes	No	Yes
R^2	0.0413	0.0982	0.0520	0.1315
F Stat	14.96*** (6, 1405)	9.93*** (59, 1405)	8.66*** (6, 757)	5.16*** (59, 757)
Observations	2015	2015	1103	1103

#Heteroskedasticity-robust standard errors (clustered at the region and sector level) are shown in parentheses.
#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001

Similar results were also found for the most and least complex regions, as relatedness density and complexity presented the same sign and the former was statistically significant only for the least complex regions, while the latter was statistically significant in both cases. Results are shown in Table C4.

Table C4 – Entry model for technologies in most and least complex regions (2006–2019)

	50% most complex regions		50% least complex regions	
Constant	-0.0387468 (0.1004796)	-0.1080021 (0.1975799)	-0.0625592 (0.0611734)	-0.921888 (0.5252149)
Relatedness density (t-1)	0.0003752 (0.0010185)	0.0001561 (0.0010867)	0.002493** (0.0007274)	0.0023916** (0.0007468)
Complexity (t-1)	0.0029814** (0.0008897)	0.0031684** (0.0009169)	-0.0019004** (0.0005759)	-0.0017647** (0.0005765)
Population density (t-1)	-0.0000285 (0.0000485)	0.0007204 (0.0020919)	0.0000106 (0.0000408)	0.0008779 (0.0016963)
GDP per capita (t-1)	-3.09e-06 (1.85e-06)	0.000012 (0.0000185)	1.71e-06 (1.00e-06)	0.0000292 (0.000019)
Size of technology (t-1)	0.0043101* (0.0018928)	0.0046092* (0.0019454)	0.0062993*** (0.0011562)	0.0063895*** (0.0011241)
Stock of regional knowledge (t-1)	0.0050669* (0.0024003)	-0.0045618 (0.0128164)	0.0040728* (0.0017123)	0.0482186 (0.041521)
Region fixed-effects	No	Yes	No	Yes
Time fixed-effects	No	Yes	No	Yes
R^2	0.0282	0.0475	0.0781	0.1520
F Stat	3.39** (6, 447)	1.61* (23, 447)	17.98*** (6, 957)	8.65*** (41, 957)
Observations	708	708	1307	1307

#Heteroskedasticity-robust standard errors (clustered at the region and sector level) are shown in parentheses.
#Coefficients are statistically significant at the *p<0.05, **p<0.01, ***p<0.001