

Relatedness and regional economic complexity: Good news for some, bad news for others

Relatedness
and regional
economic
complexity

Arthur Ribeiro Queiroz
UFMG, Belo Horizonte, Brazil

João Prates Romero
Department of Economics, UFMG, Belo Horizonte, Brazil, and

Elton Eduardo Freitas
Department of Economics, UFS, Aracaju, Brazil

Received 15 November 2023
Revised 1 February 2024
Accepted 1 February 2024

Abstract

Purpose – This article aims to evaluate the entry and exit of companies from local productive structures, with a specific focus on the sectoral complexity of these activities and the complexity of these portfolios. The study focuses on empirically demonstrating the thesis that related economic diversification exacerbates the development gap between more and less complex regions.

Design/methodology/approach – The article uses indicators formulated by the economic complexity approach. They allow a relevant descriptive analysis of the economic diversification process in Brazilian micro-regions and provide the foundation for the econometric tests conducted. Through three distinct estimation strategies (OLS, logit, probit), the influence of complexity and relatedness on the entry and exit events of firms from local portfolios is tested.

Findings – In all estimated models, the stronger relationship between an activity and a portfolio significantly increases its probability of entering the productive structure and, at the same time, acts as a significant factor in preventing its exit. Furthermore, the results reveal that the complexity of a sector reduces the probability of its specialization in less complex regions while increasing it in more complex regions. On the other hand, sectoral complexity significantly increases the probability of a sector leaving less complex local structures but has no significant effect in highly complex regions.

Research limitations/implications – Due to the data used, the indicators are calculated considering only formal job numbers. Additionally, the tests do not detect the influence of spatial issues. These limitations should be addressed by future research.

Practical implications – The article characterizes a prevailing process of uneven development among Brazilian regions and brings relevant implications, primarily for policymakers. Specifically, for less complex regions, policies should focus on creating opportunities to improve their diversification capabilities in complex sectors that are not too distant from their portfolios.

Originality/value – The article makes an original contribution by proposing an evaluation of regional diversification in Brazil with a focus on complexity, introducing a more detailed differentiation of regions based on their complexity levels and examining the impact of sectoral complexity on diversification patterns within each group.

Keywords Economic complexity, Economic diversification, Relatedness, Regional development

Paper type Research paper

JEL Classification — O38, R11, R12, R58

© Arthur Ribeiro Queiroz, João Prates Romero and Elton Eduardo Freitas. Published in *Economia*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The authors would like to thank Dominik Hartmann for his helpful comments on a previous version of this paper, as well as the comments of an anonymous referee. The usual disclaimer applies. Research grants from the National Council for Scientific and Technological Development (CNPq), numbered 408333/2023-5 and 313755/2021-3, are also gratefully acknowledged.



1. Introduction

[Hausmann and Klinger \(2007\)](#), [Hidalgo, Klinger, Barabási, and Hausmann \(2007\)](#) claim that an economy's specialization in a certain product significantly affects its future performance. This is due to the fact that economies possess different capabilities that either enable or restrict their competitiveness in producing a particular product. That is, economies with more (or less) diversified capabilities are closer (or further) to acquiring competitiveness in new sectors. Additionally, at the regional level, there is also evidence that complexity stimulates economic growth and employment ([Romero *et al.*, 2022](#)).

Empirical evidence supports the notion that economic diversification tends to occur towards sectors that share similar capabilities to be produced, as summarized by the principle of relatedness ([Hidalgo *et al.*, 2018](#)). Studies examining regional growth and the entry and exit of activities from local portfolios have demonstrated that diversification patterns in Dutch ([Frenken, Van Oort, & Verburg 2007](#)), Italian ([Boschma & Iammarino, 2009](#)), Swedish ([Neffke, Henning, & Boschma, 2011](#)), American ([Boschma, Balland, & Kogler, 2015](#); [Essletzbichler, 2013](#)) and Brazilian ([Freitas, 2019](#), [Françoso, Boschma, & Vonortas, 2022](#)) regions follow a similar trajectory. In other words, regional economic diversification tends to prioritize sectors that are already related to the existing productive structure.

However, this process reinforces the existence of a diverging pattern of diversification among regions, as observed by [Pinheiro *et al.* \(2022\)](#) in the context of European regions. Specifically, while more complex regions have the potential to diversify into more complex activities, less complex regions face significant obstacles in achieving such diversification. This structural challenge hinders their development since complex sectors offer greater economic benefits. Consequently, although diversification is driven by related sectors, it can perpetuate economic disparities between regions, indicating that relatedness can be a good news for some and a bad news for others.

The existing literature in this area is still limited, and this study aims to contribute to this theoretical field. Previous analyses by [Freitas \(2019\)](#) and [Françoso *et al.* \(2022\)](#) had a broader target, primarily assessing the influence of relatedness on regional diversification in Brazil. However, complexity emerges as a key structural determinant of diversification in these regions. Therefore, this paper aims to evaluate the entry and exit of firms in local portfolios, with a specific focus on the sectoral complexity of these activities and on the complexity of these portfolios. While [Freitas \(2019\)](#) focused on examining differences in the influence of complexity only for the most complex regions, and [Françoso *et al.* \(2022\)](#) focused solely on the influence on the probability of entry, the primary contribution of this article is to emphasize the effect of complexity on regional diversification.

To achieve this objective, this study utilizes formal employment data in productive activities across micro-regions in Brazil from 2009 to 2019. Drawing on the complexity and density indicators formulated by [Hidalgo *et al.* \(2007\)](#), we examine the influence of these two variables on the likelihood of entering or exiting productive activities within local portfolios. In addition to examining whether the principle of relatedness ([Hidalgo *et al.*, 2018](#)) holds true, as in [Freitas \(2019\)](#) and [Françoso *et al.* \(2022\)](#), where density promotes sector entry and hinders exit, our main focus is on assessing how sectoral complexity influences these probabilities, while also considering the level of regional complexity. For this, we propose a categorization of regions based on the economic complexity index (ECI): low, medium-low, medium-high and high.

These complexity-based regional groups enable the examination of previously untested hypotheses. It is assumed that sectoral complexity reduces the likelihood of new activity entry in less complex regions, while increasing it in more complex regions. Conversely, complexity raises the probability of activity exit in less complex regions but does not impact sector exit in already complex regions. By employing three different model specifications (OLS, logit and probit), the econometric results confirm the assumed hypotheses and highlight that the effect of sectoral complexity is more pronounced on the probability of

exiting sectors. Moreover, investigating the effects of product complexity and relatedness for regions at different complexity levels helps also understanding the different diversification strategies to be pursued in regions with different development levels, which is also an area still relative unexplored in the literature.

Finally, the analysis is organized as follows: [Section 2](#) discusses the literature on the subject, bringing relevant and similar contributions applied to different regions across the world. [Section 3](#) demonstrates the indicators, database and econometric specifications used. [Section 4](#) is a brief descriptive analysis. [Section 5](#) brings the results of the econometric tests and [section 6](#) ends with the final considerations.

2. Relatedness: Review of the empirical literature

The evolutionary perspective on economic change has introduced valuable concepts that enhance the understanding of the development process ([Nelson & Winter, 1982](#)). Extending this viewpoint to address the questions of economic geography has enabled the theoretical and empirical construction necessary to comprehend the evolution of regional economies and, more specifically, the determinants of their productive diversification ([Frenken *et al.*, 2007](#); [Neffke *et al.*, 2011](#); [Boschma *et al.*, 2015](#); [Françoso *et al.*, 2022](#)). Therefore, this section aims to discuss the evolution of empirical literature that evaluates regional economic development, with a focus on describing and relating how these studies explain the related diversification process.

The work of [Neffke *et al.* \(2011\)](#) marks the beginning of a series of regional studies focused on analyzing the entry, maintenance and exit of firms based on the proximity of productive structures to sectors. The authors examine 70 Swedish regions from 1969 to 1994 and investigate the influence of the number of closely related industries in a region on the probability of entry, maintenance or exit of specific sectors in the local economy. They employ three groups of estimations: models to assess (1) the probability of entry, (2) maintenance and (3) exit. The dependent variables are represented by dummy variables indicating the occurrence of these factors for each region-industry pair within five-year intervals during the specified period. The estimations utilize ordinary least squares (OLS), probit and logit models. The findings align with the assumed hypotheses, indicating that industries technologically related to existing industries are more likely to enter and persist in the regional portfolio, while those on the technological periphery are more prone to exit.

A similar analysis was conducted by [Essletzbichler \(2013\)](#) using data from 360 US metropolitan areas. Essletzbichler examined the effect of proximity on the probability of firm entry, maintenance and exit. However, there are two notable differences compared to the earlier work. First, the measurement of the proximity indicator differs. While [Neffke *et al.* \(2011\)](#) determined relatedness based on the occurrence of products from distinct industries in manufacturing plant portfolios, [Essletzbichler \(2013\)](#) measured relatedness by analyzing the intensity of flows between pairs of industries using input-output relations. The results of the analysis indicate that an industry's proximity to the regional portfolio increases the odds of membership by 6.9% and the odds of entry by 3.7%, while decreasing the odds of exit by 3.1% per additional link.

[Rigby \(2013\)](#) utilizes US patent data spanning from 1975 to 2005 to examine the impact of proximity on the likelihood of entering and exiting technology classes within cities' patent networks. The author investigates the effect of time-lagged proximity, measured by the degree of technological relatedness, on these probabilities. Linear probability models and conditional logit models are estimated using maximum likelihood techniques, with fixed effects incorporated for cities and technology classes. According to the linear probability model, increasing proximity by 1 unit for a technology in which a city has no specialization

enhances the probability of developing specialization in that area by 0.7%, while increasing proximity by 1 unit decreases the probability of developing specialization by 2.72%.

[Boschma et al. \(2015\)](#) conducted a similar analysis to [Rigby \(2013\)](#) using US patent data. However, they employed different variables and a distinct model specification. To examine whether cities diversified into related sectors, the authors accounted for potential omitted variable bias by including city characteristics and technology class variables as controls. City-level characteristics considered included employment information, population density, inventive capacity (inventors-to-employees ratio), technological specialization, growth in the number of inventors and income per employee. Technology-level variables used as controls included the number of inventors in the class, technological concentration, growth in knowledge production and a measure of patent age. The authors then tested the impact of relatedness on the entry and exit of technologies in US cities, focusing on a 5-year window within this period. They employed OLS models with fixed effects for cities, technologies, and years. To ensure the robustness of the OLS model results, alternative methods such as probit and logit were employed. The results consistently showed that relatedness density had a significant statistical and economic effect on diversification across all model specifications.

In the study conducted by [Boschma, Heimeriks, and Balland \(2014\)](#), the authors applied a similar analytical approach to analyze the influence of scientific relatedness on the emergence or disappearance of biotech research topics in the scientific portfolios of cities worldwide. The relatedness indicator was calculated based on the co-occurrence of topics in journal articles. The model specification resembled that of [Neffke et al. \(2011\)](#). As control variables, the number of publications at the city and topic levels was used. The findings indicated that new scientific topics in biotech tend to emerge in cities with existing related scientific fields. On the other hand, loosely related topics are more likely to disappear from a city's scientific portfolio.

This method was also used to assess regional diversification in Brazil. [Freitas \(2019\)](#) used employment data from Brazilian microregions to investigate regional diversification in Brazil. The study focused on the impact of density and complexity on the entry, maintenance and exit of sectors in the regions' productive structure. By incorporating economic complexity indicators into the analysis, the author hypothesized that regions would be less inclined to develop new specializations in less related and more complex activities. Using OLS, logit and probit models with fixed effects for region, productive activity and period, the study examined three 5-year windows between 2006 and 2016. The findings confirmed the hypotheses, indicating that proximity to the local productive structure increased the likelihood of sectors remaining or entering the current portfolio, while regions demonstrated reduced propensity for developing new specializations in more complex activities.

[Françoso et al. \(2022\)](#) conducted a study utilizing employment and patent data from Brazilian meso-regions to examine the impact of relatedness and complexity on regional diversification in Brazil. The authors focused on the probability of new sector entry into the local economies' productive portfolios between 2006 and 2019. To mitigate potential omitted variable bias, control variables such as population density, gross domestic product (GDP) per capita, and proxies for sector size and region diversity were employed. The study also compared the OLS results of two different samples, namely the 50% more complex and 50% less complex regions, to explore potential variations in the role of these variables. Across both employment and patent datasets, the findings demonstrated that regions tend to diversify into more related sectors, and higher levels of complexity generally reduce the probability of new sector entry. However, the relationship between complexity and sector entry is reversed in the highly complex region sample, indicating that higher complexity in such regions may actually increase the probability of new sector entry.

The inclusion of complexity indicators in the analysis of regional diversification in Brazil yields noteworthy findings that warrant further attention. In the study by [Freitas \(2019\)](#), focusing on the top 25% of regions with higher complexity, it is observed that the level of

complexity in a sector has a positive impact on the specialization of new economic activities. However, this coefficient also has a positive effect on explaining sector exit probability and a negative effect on explaining maintenance. The hypothesis put forth by the author suggests that complexity facilitates access to more complex activities but does not alleviate the "trap" of low complexity, as the probability of a sector remaining in the local structure is inversely proportional to its complexity even among the most complex regions. As previously mentioned, similar results were found by [Françoso *et al.* \(2022\)](#) when comparing the influence of complexity on activity and technological class entry in regions with varying levels of complexity.

These findings shed light on an aspect that has been largely overlooked in previous studies. The influence of complexity and relatedness in shaping regional diversification in Brazil appears to perpetuate a growing economic disparity between regions. The reversal of the complexity coefficient's sign, depending on the region, highlights an ongoing worsening of economic inequality among regions ([Pinheiro *et al.*, 2022](#); [Hartmann & Pinheiro, 2022](#)). However, this structural characteristic of the Brazilian regional diversification process was not the primary focus of the analyses conducted by [Freitas \(2019\)](#) and [Françoso *et al.* \(2022\)](#). The former examined only the diversification of the most complex microregions, while the latter only assessed the effect of complexity on the probability of entry of new productive activities. Therefore, it is crucial to investigate how the process of related diversification unfolds in Brazil, under the influence of complexity, to determine the extent of this uneven development.

The findings presented by [Pinheiro *et al.* \(2022\)](#) regarding European regions highlight a feedback loop of inequality between regions. Advanced economies tend to specialize in related high-complexity activities, while lagging and less complex regions concentrate on related low-complexity activities. The partial results mentioned by [Freitas \(2019\)](#) and [Françoso *et al.* \(2022\)](#) provide evidence of a similar pattern in Brazil, suggesting that relatedness can have positive implications for certain regions while potentially exacerbating challenges for others. This observation supports the thesis that relatedness is good news for some regions and bad news for others.

Indeed, the inability of regions to diversify into more complex activities represents a structural challenge for their development. [Hidalgo and Hausmann, \(2009\)](#) argue that engagement in complex productive sectors brings substantial economic benefits to regions due to the combination of various resources that are difficult to acquire and replicate. This creates a competitive advantage for the region, which can persist over time. In contrast, less complex activities are easier to imitate and can disperse quickly, offering less economic value and limited potential for competitive advantage. Moreover, the significance of regional complexity is empirically acknowledged and is associated with greater future GDP and employment growth ([Romero *et al.*, 2022](#)). Hence, a diversification strategy centered on highly complex activities holds greater economic benefits for regions.

Therefore, understanding the process of regional diversification in Brazil from this perspective is crucial, as the country is characterized by a structural malformation in terms of regional inequalities, extensively explored in the literature on the subject. [Silva \(2017\)](#) emphasizes that this structural issue manifests in the notion that, irrespective of the socioeconomic indicator used to gauge regional inequality, the results consistently exhibit the same pattern. The North and Northeast regions consistently show the poorest rates, while the South and Southeast regions stand out with the most favorable averages. The differences between regions play a significant role in shaping their diversification trajectories and can contribute to divergent development outcomes.

Furthermore, a specific segment of the literature delves into the convergence among municipalities and regions to scrutinize the dynamics of regional inequalities. [Vreyer and Spielvogel \(2009\)](#) estimated the speed of per capita income convergence in Brazilian

municipalities from 1970 to 1996. Their findings underscored the lack of convergence and spatial dependence among municipalities, elucidating the persistence of inequalities and the clustering of impoverished areas in less developed regions of the country. However, more recent studies, such as those by Neto (2014) and Magalhaes and Alves (2021), suggest an improvement in the scenario, particularly from the 2000s onward. Nevertheless, Magalhaes and Alves (2021) contend that this improvement does not translate into low levels of regional inequality, highlighting the enduring structural challenge of fostering sustained and less unequal national development.

When examining the panorama of Brazilian regional complexity, studies align closely with anticipated patterns. Freitas and Paiva (2016) characterize regional inequality, emphasizing its manifestation in terms of complexity. The authors point out a concentration of poles of diversity and sophistication in exports solely within the South and Southeast regions, without discernible signs of improvement in the assessed period for other regions. In a separate study, Rezende *et al.* (2023) analyze the distribution of employment across Brazilian states from 2006 to 2020. They underscore a noteworthy concentration of jobs in highly complex activities exclusively within the South and Southeast regions, highlighting an ongoing process of regressive specialization across the country.

Given this Brazilian context, the challenges faced by less complex regions in diversifying into activities that require less common capabilities should be addressed through targeted policies aimed at avoiding a potential low-complexity trap. However, to design effective policies, it is essential to examine how this logic manifests itself in Brazil. This requires a comprehensive analysis of regional dynamics, economic capabilities and the interplay between complexity, relatedness and regional development policies in order to foster inclusive and sustainable diversification across all regions in Brazil.

In this context, some hypotheses will be tested for this evaluation:

- H1.* The productive activities that enter the portfolio of a region are more related to other activities already produced in that location.
- H2.* A region is more likely to cease specialization in a particular activity when it is less related to the other activities within the local productive structure.
- H3.* Highly complex regions possess sufficient capabilities, such that increasing sector complexity enhances the probability of sector entry into their portfolios, while it has little influence on the probability of sector exit.
- H4.* Among less complex regions, greater sector complexity decreases the probability of sector entry and simultaneously acts as a significant factor driving sector exit.

The first two hypotheses, which are based on the principle of relatedness, have already been investigated in previous studies, including those conducted on regions in Brazil. Françaço *et al.* (2022) specifically examined the probability of entry for new activities and found a positive effect of relatedness, supporting Hypothesis 1. Freitas (2019) analyzed both the probability of entry and exit and found positive effects of relatedness on entry (Hypothesis 1) and negative effects on exit (Hypothesis 2). However, in addition to attesting the principle of relatedness, the subsequent hypotheses in this study aim to shed light on the persistent divergent pattern of diversification among Brazilian regions and provide insights into why relatedness is only beneficial to some regions.

In this case, hypotheses 3 and 4 have not been fully tested together in previous studies and represent the main contribution of this paper. Françaço *et al.* (2022) partially examined Hypotheses 3 and 4, finding a positive influence of sectoral complexity on the probability of entry of new activities in regions with complexity above the median and a negative influence in regions with complexity below the median. However, they did not evaluate the effect of

complexity on the probability of exiting activities from the portfolio, which is not sufficient to fully confirm [Hypotheses 3](#) and [4](#). On the other hand, [Freitas \(2019\)](#) only tested [Hypothesis 3](#) by focusing on the most complex regions (4th quartile of ECI). Although he found positive effects of sectoral complexity on the probability of entry and exit of firms, the analysis is not sufficient to confirm [Hypothesis 4](#) as it did not include less complex regions. The following section describes the methods and data that will be used to test these assumptions.

3. Data and method

3.1 Complexity measures

[Hidalgo et al. \(2007\)](#) utilized international trade data as the basis for their methodological approach, which builds upon the concept of revealed comparative advantage (RCA) introduced by [Balassa \(1965\)](#). The RCA index serves as a criterion for identifying specialization in a particular economic activity by comparing the share of that activity in the local economy to its total share in the overall economy. If the numerator (share in the local economy) is greater than the denominator (share in the overall economy), it indicates that the country or region has a competitive advantage in that sector. The RCA index shares the same conceptual framework as the location quotient (LQ) used in regional literature. The formal expression of the RCA index is presented below:

$$RCA_{ij} = \frac{X_{ij}}{\sum_i X_{ij}} \div \frac{\sum_j X_{ij}}{\sum_i \sum_j X_{ij}} \quad (1)$$

where X_{ij} represents the quantity of product i exported by country j . Therefore, if the calculation of RCA yields a value equal to or greater than 1, it indicates that country j competitively produces product i in comparison to other countries. Conversely, if the resulting index is less than 1, it implies that product i does not play a significant role in the analyzed market.

Based on this, [Hidalgo and Hausmann \(2009\)](#) propose a methodology for measuring the internal productive capacities of economies, which explains the differences in growth between countries. They analyze foreign trade data using a bipartite network approach, where countries are connected to the products they export, allowing the measurement of the complexity level of the capabilities concentrated in these economies. This measurement is based on two indicators that quantify the sophistication of products and the diversification of countries. Formally:

$$D_j = k_{c,0} = \sum_i M_{ij} \quad (2)$$

$$U_i = k_{i,0} = \sum_j M_{ij} \quad (3)$$

The quantity of goods exported with RCA serves as an indicator of the diversification of countries (D_j), while the number of countries that export a particular product with RCA reflects the ubiquity of that product (U_j). In this framework, complexity is measured based on both diversification and ubiquity. Hence, a complex country or product is characterized by high diversification and low ubiquity. The binary matrix (M_{ij}) is used to represent the sectors in which countries possess RCA, taking the value of 1 when RCA exists and 0 otherwise.

To summarize the complexity of countries and products, [Hidalgo and Hausmann \(2009\)](#) employ iterated combinations of the two indicators. These combinations are designed to weigh the characteristics when one criterion alone is insufficient to determine high or low complexity. For instance, countries with high diversification but concentrated in the

production of highly ubiquitous goods are considered less complex. Similarly, products that are not ubiquitous but produced by countries with limited diversification are also deemed less complex.

Therefore, the complexity measure depends on the eigenvector of the iteration matrix associated with its second largest eigenvalue, which captures most of the variance in the original data. Consequently, the ECI and the product complexity index (PCI) are derived from the same operations in opposite ways and are formally defined as [1]:

$$ECI = \frac{\vec{K} - \langle \vec{K} \rangle}{stdev(\vec{K})} \quad (4)$$

$$PCI = \frac{\vec{Q} - \langle \vec{Q} \rangle}{stdev(\vec{Q})} \quad (5)$$

where K and Q are the eigenvectors associated with the second largest eigenvalue, the operator $\langle \rangle$ denotes the mean, and $stdev$ represents the standard deviation.

3.2 Relatedness measure

Hidalgo *et al.* (2007) introduced the concept of proximity, which plays a crucial role in understanding the relationship between economic activities. In addition, this concept also extends to measuring the proximity or distance between the productive structures of different locations and specific goods. The framework developed by Hidalgo *et al.* (2007) centers around quantifying relatedness by examining the likelihood of two products being exported together by countries. If two goods are frequently co-exported, it implies that they share common production capabilities and are therefore related. The proximity between each pair of products is then determined by taking the minimum value among the pairwise conditional probabilities of locations that competitively produce product i , given that they also competitively produce product f . Formally, this concept can be expressed as follows:

$$\phi_{i,f} = \min\{P(RCA_i|RCA_f), P(RCA_f|RCA_i)\} \quad (6)$$

In this expression, for a location j :

$$RCA_{i,j} = \begin{cases} 1, & \text{if } RCA_{i,j} \geq 1 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

However, proximity reveals relatedness only between products. To understand the influence of relatedness in the process of productive diversification in countries and regions, an indicator capable of measuring the distance between the portfolio of an economy and a given product is needed. This indicator was also formulated by Hidalgo *et al.* (2007) and is called density. Density measures the distance between a given good and the productive portfolio of a location. This index also represents the difficulty for a location to specialize in a sector since the further away the product is from the local portfolio, the lower the chances of having common capabilities for its production. Therefore, we call this indicator relatedness density, as defined by Boschma *et al.* (2014):

$$RelatednessDensity_{i,j} = \frac{\sum_{f \in j, f \neq i} \phi_{i,f}}{\sum_{f \neq i} \phi_{i,f}} \times 100 \quad (8)$$

Equation (8) demonstrates the relatedness density between a product i and the productive structure of a given country j . The indicator is the sum of the proximities between product i

and the other goods that country j has RCA, weighted by the sum of the proximities between this product and all other goods. It represents the weighted proportion of goods related to product i that are competitively produced by country j . This indicator varies between 0% and 100%. A density of 0% for a given good i and country j means that there are no other related products in that country's portfolio. On the other hand, a density of 100% means that all goods related to product i are competitively produced by country j .

3.3 Data

To assess the diversification process of Brazilian microregions, our main data source will be employment data in economic-productive activities. Employment data has been widely utilized for subnational analyses as it provides more up-to-date information, covers the entire territorial dimension and offers a high level of specification (Freitas, 2019; Franoso *et al.*, 2022; Romero *et al.*, 2022). Unlike foreign trade data, which was utilized by Hidalgo *et al.* (2007) and Hidalgo and Hausmann (2009), employment data is more suitable for regional analysis in Brazil due to the large number of municipalities that do not engage in exporting or importing activities, thereby lacking relevant information. Moreover, considering the significant weight of the domestic market in the Brazilian economy, employment data provides a comprehensive perspective.

Therefore, in order to measure the aforementioned indicators, we utilized the LQ instead of the RCA and adapted the concept of proximity based on the co-location of productive activities among Brazilian microregions. Formally, expressions (1) and (6) are constructed as follows:

$$LQ_{i,j} = \frac{Emp_{i,j}}{\sum_i Emp_{i,j}} \div \frac{\sum_j Emp_{i,j}}{\sum_j \sum_i Emp_{i,j}} \quad (9)$$

$$\phi_{i,f} = \min\{P(LQ_i|LQ_f), P(LQ_f|LQ_i)\} \quad (10)$$

In equation (9), the quantity exported is replaced by employment in productive activity i and microregion j . In equation (10), proximity is calculated based on the concept of co-location between activities, which is determined by the number of locations where both activities are competitively and jointly produced. Thus, the proximity between activities i and f is defined as the minimum conditional probability that the LQ is greater than 1 in one activity, given that the LQ is greater than 1 in the other activity. This adaptation enables the measurement of indicators for Brazilian microregions and, consequently, facilitates the analysis of the diversification proposed in this paper.

The main data source for this study is the annual social information report (RAIS), organized by the Ministry of Labor and Employment. This database contains mandatory administrative records for formal employment establishments in Brazil. From the RAIS, we extracted information on the number of formal jobs by municipality (and region) and by economic activity sector. Economic activities were grouped based on the 6-digit class of the National Classification of Economic Activities (CNAE) proposed by the Brazilian Institute of Geography and Statistics (IBGE). The chosen territorial unit for analysis is the microregions [2] (IBGE, 1990). This approach allowed us to organize a database comprising 558 Brazilian micro-regions and 670 productive activities classified according to the CNAE class. Furthermore, it is essential to note that the results to be presented must be interpreted with the awareness that the data exclusively reflects the formal job market. This limitation is a crucial consideration for future research to address, since estimates indicate that about 40% of the employment in Brazil is informal.

3.4 Econometric specifications

The model specification follows the pattern tested by [Neffke et al. \(2011\)](#) and replicated in several other articles. Hence, we opted to assess the impact of relatedness density and sector complexity on the probability of entry and exit of productive activities in the portfolios of Brazilian regions. To do so, we will use binary dependent variables that indicate the occurrence of entry or exit for a specific activity in a given region.

The variable *Entry* is defined as 1 if a micro-region j is not specialized in economic activity i at time t ($LQ < 1$), but becomes specialized at $t + 5$ ($LQ \geq 1$). It takes the value 0 when the microregion was not specialized at time t and also does not become specialized at time $t + 5$. So, it considers only the subset of activities that were not competitively produced by the micro-regions at time t ($LQ < 1$). On the other hand, the variable *Exit* follows the opposite logic. It is assigned the value 1 when microregion j was specialized in activity i at time t ($LQ \geq 1$) but ceases to be specialized at time $t + 5$ ($LQ < 1$). It is assigned the value 0 when the microregion was specialized at time t and continues to be specialized at time $t + 5$. Therefore, for the "Exit" variable, the observations are limited to cases where the activity was competitively produced at time t ($LQ \geq 1$). Formally, the definitions are as follows:

$$Entry_{j,i,t} = I(i \notin PF(j, t) \cap i \in PF(j, t + 5)) \quad (11)$$

$$Exit_{j,i,t} = I(i \in PF(j, t) \cap i \notin PF(j, t + 5)) \quad (12)$$

The choice of 5-year intervals follows the approach adopted by [Boschma et al. \(2014, 2015\)](#) and [Freitas \(2019\)](#). We selected two 5-year periods between 2009 and 2019 (2009–2014 and 2014–2019), resulting in a balanced and complete panel with 1,121,580 observations. However, for the estimation of the entry and exit models, the panel is further reduced. For the dummy variable *Entry*, only activities that have the potential to enter the portfolio in the subsequent period are considered. This means that the LQ must be less than 1 in the initial periods (2009 or 2014). As a result, the subsample used for the entry model comprises 647,801 observations. For the *Exit* dummy variable, we consider only activities that could potentially leave the portfolio of microregions in the following period ($LQ \geq 1$ in 2009 or 2014), resulting in a subsample of 99,919 observations.

The specification of the models is as follows:

$$Entry_{j,i,t} = \beta_1 RelatednessDensity_{j,i,t-5} + \beta_2 PCI_{i,t-5} + \beta_3 Regions_{j,t-5} + \beta_4 Activities_{i,t-5} + \phi_j + \psi_i + \varepsilon_{j,i,t} \quad (13)$$

$$Exit_{j,i,t} = \beta_1 RelatednessDensity_{j,i,t-5} + \beta_2 PCI_{i,t-5} + \beta_3 Regions_{j,t-5} + \beta_4 Activities_{i,t-5} + \phi_j + \psi_i + \varepsilon_{j,i,t} \quad (14)$$

where $Regions_{j,t-5}$ is the vector of variables used to control for observable characteristics that vary over time in Brazilian microregions. Similarly, $Activities_{i,t-5}$ is a vector of variables that summarize the characteristics of productive activities. These variables are presented in [Table 1](#). The fixed effects are represented by ϕ_j for regions and ψ_i for activities. Finally, $\varepsilon_{j,i,t}$ represents the residuals.

The main independent variables for the analyses conducted in this article are relatedness density and PCI. However, we also include additional variables to control for specific characteristics of regions and activities. The following table summarizes the control variables used.

In addition to controlling for structural characteristics such as per capita GDP, population, human capital, productivity, presence of entrepreneurial incentives and average sector size, we also include variables to capture the effects of local economy diversity (Diversity) and sector spatial concentration (coefficient of localization (CL)). The inclusion of the diversity

Variables	Description	Operation	Source	Vector
GDPpc	Gross Domestic Product per capita	$\frac{\text{Gross Domestic Product}_j}{\text{Population}_j}$	IBGE	Regions_j
Population	Micro-regions population	Population_j	IBGE	Regions_j
Region productivity	Average between micro-region's salary mass and the number of formal jobs	$\frac{\text{Salary Mass}_j}{\text{Employment}_j}$	RAIS	Regions_j
Region HC	Percentage of workers with at least an incomplete undergraduate degree in the microregion	$\frac{\text{High Skilled Employment}_j}{\text{Employment}_j} \times 100$	RAIS	Regions_j
Incentives	Percentage of municipalities with incentives to attract entrepreneurial activities	$\frac{\text{Municipalities with Incentives}}{\text{Total of Municipalities}} \times 100$	MUNIC ^a	Regions_j
Credit	Volume of credit operations per capita	$\frac{\text{Volume of credit operations}_i}{\text{Population}_i}$	ESTBAN ^b	Regions_j
Diversity	Diversity measured by Shannon (1948)'s index	$e^{-\sum_{i=1}^I \frac{\text{Employment}_{i,j}}{\text{Employment}_j} \ln \left(\frac{\text{Employment}_{i,j}}{\text{Employment}_j} \right)}$	RAIS	Regions_j
Sector size	Average number of employees per establishment	$\frac{\text{Employment}_i}{\text{Establishment}_i}$	RAIS	Activities_i
Sector HC	Percentage of workers with at least an incomplete undergraduate degree in the activity	$\frac{\text{High Skilled Employment}_i}{\text{Employment}_i} \times 100$	RAIS	Activities_i
Sector productivity	Average between activity's salary mass and the number of formal jobs	$\frac{\text{Salary Mass}_i}{\text{Employment}_i}$	RAIS	Activities_i
CL	Spatial concentration of activity measured by the coefficient of localization (Florence, 1948)	$\frac{1}{2} \sum_j \left \frac{\text{Employment}_{i,j}}{\text{Employment}_i} - \frac{\text{Employment}_j}{\text{Employment}} \right $	RAIS	Activities_i

Relatedness and regional economic complexity

Note(s): ^aMunicipal Basic Information Survey (MUNIC) conducted by IBGE
^bMonthly Banking Statistics by Municipality (ESTBAN) provided by the Central Bank
Source(s): Own elaboration

Table 1.
Control variables

variable aligns with the literature that aims to understand the influence of a diverse local economy in attracting new industries (Glaeser, Kallal, Scheinkman, & Shleifer, 1992, Henderson, Kuncoro, & Turner 1995). Furthermore, the CL is utilized to account for the inherent difficulty of attracting or exiting a particular sector. It is assumed that a higher spatial concentration of an activity indicates a lower probability of entry or exit.

Finally, it is important to note that regressions 13 and 14 will be estimated taking into account the complexity of each region. The micro-regions will be divided into four complexity groups, which are expected to exhibit different diversification processes according to the hypotheses. The classification criteria used will be explained and discussed in the section below.

4. Complexity groups

The hypotheses of this study will be tested by grouping the regions according to their level of complexity. The stratification will be based on the ECI value, resulting in four distinct groups of microregions:

- (1) *Low complexity*: microregions with an ECI up to 0.25.
- (2) *Medium-low complexity*: microregions with an ECI between 0.25 and 0.50.
- (3) *Medium-high complexity*: microregions with an ECI between 0.50 and 0.75.
- (4) *High complexity*: microregions with an ECI above 0.75.

The existing literature has not yet converged on a unanimous approach to classifying regions based on their complexity levels. Prevailing contributions often distinguish regions according to the distribution of the ECI, as demonstrated in the studies by Freitas (2019) and Franoso *et al.* (2022). However, this method frequently results in the grouping of regions with significantly divergent complexity levels. To address this concern, we have adopted a strategy that differentiates regions based on their specific index values. While this approach may yield groups with varying numbers of regions, it ensures a more equitable representation of complexity levels across these groups. It is essential to note, however, that this strategy relies on the establishment of arbitrary values for differentiating the ECI.

Figure 1 a, b, and c show the distribution of microregions across these complexity groups for the reference years of analysis. The configuration of the groups appears to be consistent and stable across all years. Microregions with high and medium-high complexity are primarily concentrated in the South and Southeast regions, particularly around major urban centers. Microregions with low and medium-low complexity are located in more inland regions as well as in the North, Northeast and midwest regions.

Figure 1 d and e illustrate the distribution of ECI within each complexity group, with separate graphs for entry models and exit models. As mentioned earlier, the reason for this separation is that these models represent different sets of activities. The distribution within each group reveals distinct patterns. Among microregions with low complexity, the ECI values are predominantly concentrated around 0.25, with some outliers towards the lower end of the range. For microregions with medium-low complexity, the distribution tends to be closer to 0.50. Similarly, microregions with medium-high complexity also show a distribution closer to 0.50, indicating concentration towards the lower values of the range. Finally, highly complex microregions tend to have ECI values close to the lower limit of the range, around 0.75.

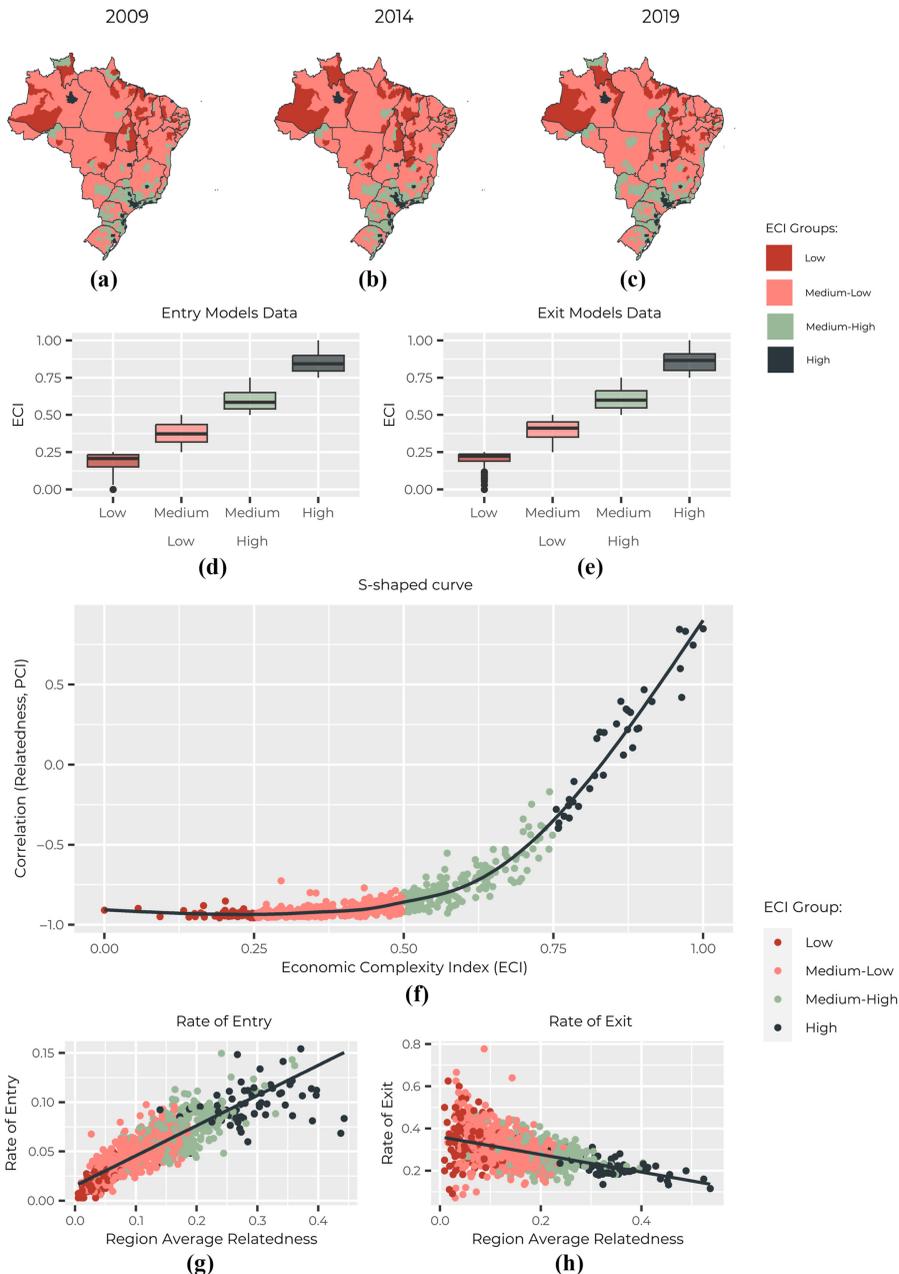
The visualization of the S-shaped curve, representing the relationship between region complexity and proximity to new complex activities, is crucial for testing the hypotheses in this study. Figure 1 f presents this curve for Brazilian microregions, which exhibits a distinct configuration compared to European regions (Pinheiro *et al.*, 2022) and other countries (Hartmann, Bezerra, Lodolo, & Pinheiro, 2020). Unlike the European context, Brazil's regional development features two primary stages. The first stage includes regions where an increase in complexity does not significantly reduce the distance from less complex productive structures (ECI between 0 and 0.50). The second stage comprises intermediate regions, where even small increments in complexity lead to substantial increases in proximity to complex activities (ECI between 0.50 and 1.00). This nuanced pattern highlights the emphasis on the intermediate stage of development in Brazil, where a significant number of complex microregions exhibit varying levels of proximity to new complex activities.

Finally, to complete the analysis, it is necessary to evaluate the dynamics of entry and exit of firms according to the previously defined complexity groups. Following Boschma *et al.* (2015), Figure 1 g and h show, respectively, the entry and exit rates of activities in the portfolio of microregions according to the average density in each one of them. In addition, the color of the dots identifies the groups to which each region belongs. Figure 1 g shows that the entry rate is well correlated with the average relatedness of the region, except that the groups of greater complexity present a greater dispersion around the line. On the other hand, for activity exit rates (Figure 1 h), there is a negative relationship with the average density of the regions, but a weaker correlation, mainly due to the less complex groups.

5. Econometric tests

Figure 2 illustrates the strength of the correlation among the dependent and independent variables used in the models. Correlations are in most cases positive, indicating some degree

Relatedness and regional economic complexity



Source(s): Own elaboration

Figure 1.
Complexity groups

of relationship, although they are generally weak. The strongest correlations are observed between diversity and relatedness (0.82) and between regional productivity and GDP per capita (0.65). This can be explained by the fact that more diversified economies possess a

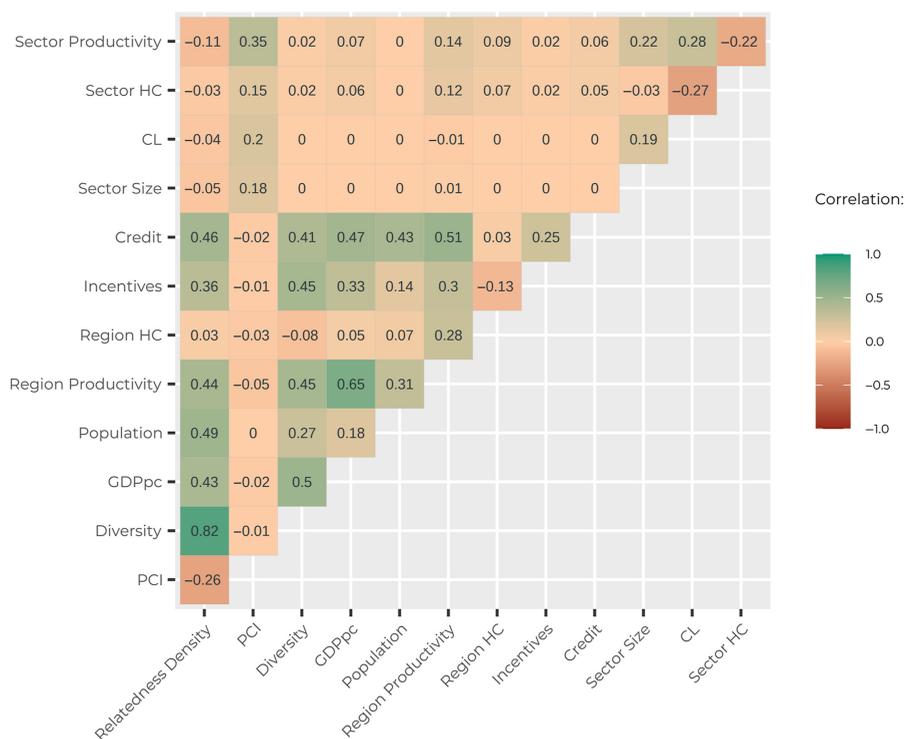


Figure 2.
Correlogram

Source(s): Own elaboration

range of capabilities that facilitate entry into new sectors. In the latter case, it is assumed that wealthier regions have higher average worker salaries, which serves as a proxy for productivity. Importantly, there is no significant collinearity among the regressors that would impede the estimation of the models.

Table 2 presents the construction of the final model, which is estimated by differentiating the regions into complexity groups. The table consists of six estimates: regression (1) measures the influence of only the main variables (relatedness density and PCI); regression (2) considers only variables controlling observable characteristics of microregions; regression (3) considers only control variables for activities; and regressions (4), (5) and (6) are the final models specified in equation (13). Moving on to Table 3, it presents the results of the final model considering the groups of microregions by complexity. Since the estimates using the three estimation strategies (OLS, logit and probit) are similar and consistent with each other, we chose to focus on presenting the results of the logit model in this section for simplicity [3]. However, the corresponding estimates using OLS and probit can be found in Annex, which ensures the robustness of the logit model. Tables 4 and 5 follow the same approach, but for the dependent variable *Exit*.

The results presented in Table 2 support a consistent narrative. While the intensity of the coefficients cannot be evaluated due to the reasons discussed earlier, comparing models (4), (5) and (6) reveals that the direction of the effects of the independent variables are consistent across all three specifications, and the same variables remain significant. Consequently, the hypothesis of the principle of relatedness (Hidalgo *et al.*, 2018) is validated, indicating a

	Dependent variable					
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	Logit (5)	Probit (6)
Relatedness density	0.409 ^{***} (0.018)			0.646 ^{***} (0.029)	9.839 ^{***} (0.626)	4.931 ^{***} (0.285)
PCI	-0.098 ^{***} (0.016)			-0.049 ^{**} (0.019)	-2.235 ^{***} (0.426)	-1.045 ^{***} (0.205)
Diversity		0.012 ^{***} (0.001)		-0.005 ^{***} (0.001)	0.055 [*] (0.031)	0.020 (0.014)
Log(GDPpc)		0.001 (0.002)		-0.002 (0.002)	-0.071 (0.024)	-0.024 (0.024)
Log(population)		0.008 ^{***} (0.001)		-0.010 ^{***} (0.002)	-0.218 ^{***} (0.039)	-0.097 ^{***} (0.018)
Region productivity		-0.011 ^{***} (0.003)		-0.004 (0.003)	-0.103 (0.091)	-0.055 (0.041)
Region HC		0.007 (0.006)		-0.026 ^{***} (0.008)	-0.495 ^{***} (0.221)	-0.251 ^{**} (0.099)
Incentives		0.0002 (0.002)		0.003 (0.002)	0.061 (0.053)	0.033 (0.024)
Credit		0.000 (0.000)		-0.000 ^{**} (0.000)	-0.000 [*] (0.000)	-0.000 [*] (0.000)
Sector size			-0.007 ^{***} (0.001)	-0.004 ^{***} (0.001)	-0.175 ^{***} (0.029)	-0.076 ^{***} (0.013)
CL			-0.090 ^{***} (0.009)	-0.084 ^{***} (0.009)	-2.047 ^{***} (0.224)	-0.982 ^{***} (0.101)
Sector HC			-0.020 [*] (0.012)	0.008 (0.011)	0.056 (0.245)	0.037 (0.114)
Sector productivity			-0.011 ^{***} (0.002)	-0.002 (0.003)	-0.143 [*] (0.076)	-0.071 ^{**} (0.035)
Constant	0.021 ^{**} (0.010)	-0.021 (0.024)	0.212 ^{***} (0.019)	0.254 ^{***} (0.031)	2.470 ^{***} (0.750)	0.893 ^{***} (0.347)
Fixed effects – UF	Yes	Yes	Yes	Yes	Yes	Yes

(continued)

Relatedness
and regional
economic
complexity

Table 2.
Emergence of new
activities in Brazilian
microregions
(2009–2019)

Table 2.

	Dependent variable					
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	Logit (5)	Probit (6)
Fixed effects – activities	Yes	Yes	Yes	Yes	Yes	Yes
Observations	647,801	647,801	647,801	647,801	647,801	647,801
R ²	0.044	0.032	0.033	0.049		
Pseudo R ²	0.213	0.214	0.214	0.212	0.12	0.12
Residual Std. Error	260.428 ^{***}	181.883 ^{***}	188.327 ^{***}	266.430 ^{***}		
F-statistic						

Note(s): ^{*} $p < 0.1$; ^{**} $p < 0.05$; ^{***} $p < 0.01$
Robust standard-errors (clustered at the micro-region and activity level) are in parentheses
Initial periods (t) are 2009 and 2014 and final (t+5) 2014 and 2019
Source(s): Own elaboration

	Dependent variable					Relatedness and regional economic complexity
	General model (1)	Low (2)	Medium-low (3)	Medium-high (4)	High (5)	
Relatedness density	9.839*** (0.626) [0.436]	23.779*** (3.271) [0.437]	11.794*** (1.087) [0.460]	13.531*** (1.256) [0.838]	16.059*** (1.427) [1.279]	
PCI	-2.235*** (0.426) [-0.099]	-6.388*** (0.839) [-0.117]	-3.584*** (0.511) [-0.140]	-0.180 (0.435) [-0.011]	0.773 (0.300) [0.062]	
Diversity	0.055* (0.031)	-0.079 (0.110)	-0.079* (0.044)	-0.542*** (0.090)	-0.581** (0.252)	
Log(GDPpc)	-0.071 (0.053)	-0.161 (0.210)	0.006 (0.046)	0.055 (0.084)	-0.445* (0.246)	
Log(population)	-0.218*** (0.039)	0.128** (0.057)	-0.030 (0.029)	-0.143*** (0.051)	-0.799*** (0.148)	
Region productivity	-0.103 (0.091)	-0.724*** (0.274)	-0.235** (0.093)	0.062 (0.150)	0.389 (0.352)	
Region HC	-0.495** (0.221)	-0.433 (0.306)	-0.499** (0.205)	-0.966** (0.415)	1.088 (1.102)	
Incentives	0.061 (0.053)	-0.204 (0.142)	0.087* (0.051)	0.109 (0.113)	0.119 (0.425)	
Credit	-0.000* (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	
Sector size	-0.175*** (0.029)	-0.187*** (0.066)	-0.196*** (0.035)	-0.161*** (0.033)	-0.161*** (0.043)	
CL	-2.047*** (0.224)	-1.671*** (0.421)	-2.091*** (0.254)	-2.366*** (0.221)	-2.576*** (1.257)	
Sector HC	0.056 (0.245)	0.135 (0.497)	-0.345 (0.306)	0.554** (0.251)	0.691** (0.348)	
Sector productivity	-0.143* (0.076)	-0.115 (0.168)	-0.135 (0.089)	-0.138* (0.076)	0.002 (0.120)	
Constant	2.470*** (0.750)	2.806 (1.748)	1.878*** (0.722)	2.296** (1.056)	7.086*** (2.416)	
Fixed effects – UF	Yes	Yes	Yes	Yes	Yes	
Fixed effects – activities	Yes	Yes	Yes	Yes	Yes	
Observations	647,801	71,305	392,864	153,505	30,127	
Pseudo R2	0.12	0.2	0.14	0.09	0.11	
Log likelihood	-112612.200	-5628.153	-60592.810	-35442.190	-8413.916	
Akaike Inf. Crit.	225476.300	11492.310	121435.600	71130.390	17049.830	

Note(s): * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$
Robust standard-errors (clustered at the microregion and activity level) are in parentheses
Initial periods (t) are 2009 and 2014 and final ($t+5$) 2014 and 2019
The values in square brackets [] represent the average marginal effects
Source(s): Own elaboration

Table 3.
Emergence of new activities – logit models

positive relationship between relatedness density and the probability of entry. Therefore, acquiring a diverse set of skills that enhances the micro-regions' capacity to specialize in other activities is crucial for regional economic diversification. As mentioned, this confirmation is also present in previous studies on Brazil. However, the analysis of the PCI's role introduces complications, as, on average, sectors with higher complexity are less likely to become specialized within a region. This suggests that the process of accumulating skills and diversifying the economic activities is not straightforward. To delve deeper into this issue,

ECON

	Dependent variable					
	Exit		Exit			
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	Logit (5)	Probit (6)
Relatedness density	-0.820*** (0.036)			-1.334*** (0.082)	-7.745*** (0.501)	-4.507*** (0.291)
PCI	0.389*** (0.045)			0.368*** (0.043)	1.848*** (0.238)	1.102*** (0.140)
Diversity		-0.015*** (0.005)		0.027*** (0.006)	0.178*** (0.030)	0.102*** (0.018)
Log(GDPpc)		0.001 (0.008)		0.005 (0.008)	0.015 (0.043)	0.012 (0.025)
Log (population)		-0.034*** (0.004)		0.023*** (0.005)	0.132*** (0.028)	0.081*** (0.017)
Region productivity		0.016 (0.012)		-0.021 (0.016)	-0.115 (0.084)	-0.066 (0.051)
Region HC		0.042 (0.033)		0.103*** (0.040)	0.579*** (0.209)	0.338*** (0.125)
Incentives		-0.001 (0.009)		-0.009 (0.011)	-0.051 (0.056)	-0.031 (0.034)
Credit		-0.000** (0.000)		0.000** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Sector size			-0.049*** (0.006)	-0.054*** (0.006)	-0.298*** (0.033)	-0.174*** (0.019)
CL			0.214*** (0.044)	0.290*** (0.039)	1.559*** (0.206)	0.932*** (0.123)
Sector HC			0.153*** (0.056)	0.112** (0.053)	0.595** (0.278)	0.361** (0.166)
Sector productivity			0.028** (0.011)	0.023* (0.013)	0.132* (0.069)	0.074* (0.041)
Constant	0.330*** (0.029)	0.594*** (0.074)	-0.012 (0.081)	-0.185** (0.093)	-3.644*** (0.501)	-2.211*** (0.298)
Fixed effects – UF	No	No	No	Yes	Yes	Yes
Fixed effects – activities	No	No	No	Yes	Yes	Yes
Observations	99,919	99,919	99,919	99,919	99,919	99,919
R ²	0.072	0.057	0.055	0.082		
Pseudo R ²					0.07	0.07
Residual Std. Error	0.431	0.435	0.435	0.429		
F-statistic	67.756***	50.585***	50.211***	71.541***		

Table 4. Exit of activities in Brazilian microregions (2009–2019)

Note(s): * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard-errors (clustered at the microregion and activity level) are in parentheses. Initial periods (t) are 2009 and 2014 and final ($t+5$) 2014 and 2019. **Source(s):** Own elaboration

further analysis considering the complexity level of microregions is warranted. This dynamic, although cited, was not the primary focus of attention by Freitas (2019) and Françoso *et al.* (2022) and will be our main contribution.

Table 3, in turn, represents the first part of the main contribution of this article. Previous studies either do not achieve such a level of disaggregation when differentiating regions by complexity (Françoso *et al.*, 2022) or solely focus on the most complex regions (Freitas, 2019). The segmented analysis based on the complexity level of regions reveals the inherent inequality in the diversification process of Brazilian microregions. Across all three estimation strategies, the relatedness density consistently shows a significant and positive effect on the

	Dependent variable					Relatedness and regional economic complexity
	General model (1)	Low (2)	Medium-low (3)	Medium-high (4)	High (5)	
Relatedness density	-7.745*** (0.501)	-11.517** (4.984)	-12.222*** (1.019)	-9.621*** (1.261)	-10.978*** (1.289)	
PCI	[-1.419] 1.848*** (0.238)	[-2.095] 7.601*** (1.407)	[-2.239] 2.687*** (0.370)	[-1.760] 0.569** (0.271)	[-1.730] -0.396 (0.309)	
Diversity	[0.338] 0.178*** (0.030)	[1.383] 0.196 (0.180)	[0.492] 0.350*** (0.047)	[0.104] 0.648*** (0.093)	[-0.062] 0.422** (0.170)	
Log(GDPpc)	0.015 (0.043)	-0.592*** (0.203)	-0.011 (0.052)	-0.001 (0.079)	0.277** (0.120)	
Log(population)	0.132*** (0.028)	-0.059 (0.108)	0.024 (0.032)	0.067 (0.050)	0.524*** (0.114)	
Region productivity	-0.115 (0.084)	1.171*** (0.369)	-0.073 (0.104)	-0.268* (0.159)	-0.853*** (0.218)	
Region HC	0.579*** (0.209)	-0.390 (0.649)	0.746*** (0.248)	0.961** (0.454)	1.621* (0.902)	
Incentives	-0.051 (0.056)	-0.030 (0.235)	-0.130* (0.072)	0.053 (0.126)	-0.292 (0.344)	
Credit	0.000*** (0.000)	-0.000 (0.00000)	0.000 (0.000)	0.000 (0.000)	0.000** (0.000)	
Sector size	-0.298*** (0.033)	-0.305*** (0.124)	-0.302*** (0.041)	-0.285*** (0.036)	-0.264*** (0.043)	
CL	1.559*** (0.206)	1.456** (0.727)	1.431*** (0.247)	1.957*** (0.222)	3.357*** (0.290)	
Sector HC	0.595** (0.278)	-0.229 (0.963)	0.622* (0.369)	0.407 (0.282)	0.838** (0.330)	
Sector productivity	0.132* (0.069)	0.388 (0.261)	0.149 (0.092)	0.096 (0.079)	0.020 (0.102)	
Constant	-3.644*** (0.501)	-10.714*** (2.753)	-2.934*** (0.665)	-3.909*** (0.984)	-5.220*** (1.984)	
Fixed effects – UF	Yes	Yes	Yes	Yes	Yes	
Fixed effects – activities	Yes	Yes	Yes	Yes	Yes	
Observations	99,919	3,065	47,326	36,775	12,753	
Pseudo R2	0.07	0.15	0.10	0.06	0.07	
Log likelihood	-54737.750	-1658.514	-25890.140	-20160.710	-6223.971	
Akaike Inf. Crit	109727.500	3535.029	52030.280	40567.430	12669.940	
Note(s): * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$						
Robust standard-errors (clustered at the microregion and activity level) are in parentheses						
Initial periods (t) are 2009 and 2014 and final ($t+5$) 2014 and 2019						
The values in square brackets [] represent the average marginal effects						
Source(s): Own elaboration						

Table 5.
Exit of activities – logit
models

probability of a new activity entering the local productive structures, irrespective of the region's complexity level (**Hypothesis 1**). However, the complexity of regional portfolios differentiates the influence of PCI on the probability of new sector emergence. In the less complex groups (low and medium-low), an increase in PCI negatively affects the likelihood of a particular activity specializing in these regions. The medium-high complexity group appears to be in a transitional position, with varying coefficient signs across models and without statistical significance. In contrast, the microregions with high complexity (high)

exhibit positive and significant coefficients for PCI in all models. This indicates that the impact of activity complexity on the probability of new sector entry is reversed, becoming positive. This pattern reinforces the thesis that relatedness is good news only for some, as only the most complex regions possess capabilities that enable production in new, more complex sectors.

In quantitative terms, the results also demonstrate economic significance. Since we cannot interpret the coefficients directly as in OLS, the values enclosed in square brackets in [Table 3](#) illustrate the impact of relatedness density and PCI on the probability of entry through the average marginal effects. A 0.1 increase in relatedness density corresponds to an approximately 4–5% increase in the probability of entry for low and medium-low complexity regions, an 8% increase for medium-high, and a 13% increase for high. As for the PCI variable, which is central to our argument, a 0.1 increase in the indicator results in a decrease of 1.2–1.4% in the probability of entry for regions with low and medium-low complexity, and an increase of 0.6% for regions with high complexity.

[Tables 4 and 5](#) present the same estimates for assessing the probability of activities exiting. Comparing models (4), (5) and (6) in [Table 4](#), the coefficients maintain the same sign, and the significant variables remain the same across the models. Once again, the principle of relatedness is supported, as the effect of relatedness density is consistently negative and significant in all estimates. This implies that a higher density decreases the likelihood of the activity ceasing to be specialized in the region. Conversely, complexity exerts a contrasting force, increasing the probability of an activity leaving. This finding aligns with the results found by [Freitas \(2019\)](#). The control variables exhibit effects opposite to those in the entry models, and generally, the same variables remain significant.

However, what is crucial here is to understand the impact of these variables while considering the differentiation of the complexity level among micro-regions. This represents the second part of our contribution, as it involves an analysis that is missing in previous works. [Freitas \(2019\)](#) examines the probability of exit only for the top 25% most complex regions, while [Françoso et al. \(2022\)](#) does not assess the removal of activities from the local portfolio. Once again, the results demonstrate an uneven diversification process, as the same pattern emerges in all three estimates. Relatedness density plays a role in reducing the probability of activity exit, regardless of the level of complexity. Nonetheless, the complexity of sectors is pivotal in increasing the likelihood of exiting the local portfolio, particularly in less complex groups (low and medium-low). On the other hand, as the complexity of microregions increases, the effect of PCI diminishes and becomes insignificant. This phenomenon highlights that less complex microregions lack the necessary skills, knowledge and capabilities to sustain complex activities within their structure, while more complex regions do not face the same challenge.

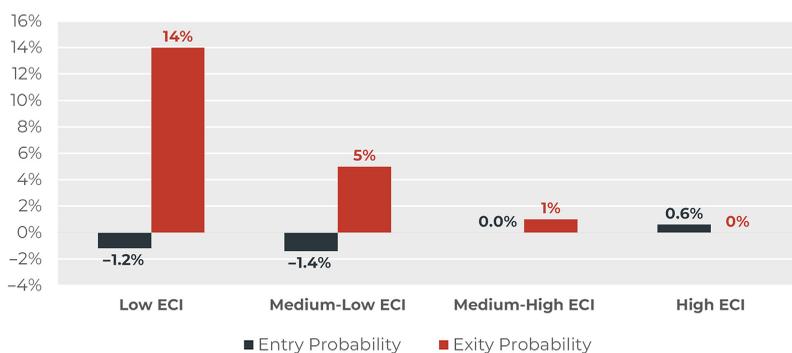
In terms of quantitative interpretation, the coefficients reveal a more pronounced effect of the main variables on the probability of exiting activities compared to the entry models. According to [Table 5](#), a 0.1 increase in relatedness density results in approximately a 21% decrease in the probability of activities leaving low-complexity local structures. This reduction is 22% for medium-low and 17% for medium-high and high complexity microregions. On the other hand, the PCI has a greater impact on the probability of activity exit than on the entry of new sectors. Among low-complexity groups, a 0.1 increase in the PCI translates to a 14% increase in the exit probability for low regions and a 5% increase for medium-low regions. However, for more complex regions, the effect diminishes to 1.0% for medium-high and becomes insignificant for high complexity micro-regions.

The econometric tests conducted in this paper validate the hypotheses proposed. Across all estimated models, relatedness density strongly promotes the entry of new sectors into local structures and simultaneously acts as a significant factor in preventing the exit of existing sectors (supporting [Hypotheses 1 and 2](#)). While these results have been empirically demonstrated before, this work contributes by delving deeper into the study within the context of activity and regional complexity that had not yet been fully studied ([Hypotheses 2 and 3](#)). Furthermore, the results reveal that the complexity of a sector reduces the probability of its specialization in less complex regions while increasing it in more complex regions. Conversely, sectoral complexity significantly raises the likelihood of a sector leaving less complex local structures but has no significant effect in high complexity regions. These findings provide evidence for [Hypotheses 3 and 4](#), supporting the notion that the process of related diversification is inherently uneven.

It is also worth noting that complexity has a stronger impact on the probability of sectoral exits compared to its effect on attracting new activities. [Figure 3](#) summarizes the percentage changes in the probability of entry and exit due to a 0.1 increase in the PCI of the targeted activity. The greater effect on exit probability at lower levels of regional complexity indicates that unrelated diversification is very risky when it comes to low complexity regions. Moreover, this means also that it is much harder to keep complex activities in least developed regions than attracting them. This finding has important policy implications, stressing the importance of additional policies in relation to the usual investment attraction strategies.

[Figure 3](#) also demonstrates that the effect of entering activities with higher complexity becomes weaker in regions at the medium-high complexity level. This suggests that this is the least risky development level to pursue unrelated diversification. The results show also that medium-low complexity regions could be the target of hybrid strategies, since the exit risk reduces considerably in relation to low complexity regions.

Furthermore, the econometric results reinforce the interpretation discussed in [Figure 1 f](#), which illustrates an exponential curve. The figure helps us understand that as regional complexity increases, there is a closer proximity to more complex sectors, but only among those with an ECI of at least 0.50. Attracting more complex sectors to less complex regions is indeed a challenging task. This is why the PCI has a negative impact on the probability of entry for new sectors and a notably positive impact on the probability of exit among low and medium-low regions. Conversely, in regions of high complexity, the PCI increases the probability of a given activity becoming specialized and does not significantly affect the probability of exit. This is because the distance from more complex sectors is much



Source(s): Own elaboration

Figure 3.
Changes in the probabilities of entry and exit in response to an increase in 0.1 in the complexity of the targeted activity

smaller in high complexity regions and decreases exponentially as regional complexity increases. The steep slope of the curve in this group indicates that the higher the regional complexity, the greater the proximity gains to more complex sectors.

6. Concluding remarks

This article introduced new elements for understanding the inherent inequality in the diversification process. An analysis under the prism of economic complexity highlights the challenges that less complex regions face in diversifying their economies, as well as the ability of more complex regions to maintain a productive structure that attracts and retains more complex activities.

The descriptive analysis of formal employment data from 568 Brazilian microregions yielded significant findings. Firstly, complexity in Brazil is regionally concentrated, with the highest rates observed in regions consisting of state capitals. Furthermore, in contrast to its application in other regions, the S-shaped curve actually takes on an exponential form for Brazilian microregions. Instead of development occurring in two distinct stages, as argued by [Pinheiro *et al.* \(2022\)](#), the composition in Brazil consists of one extreme stage represented by less complex regions and a transitional stage represented by more complex regions. While the former are trapped in a cycle where increases in complexity do not result in a closer proximity to more complex products, the latter experience that even small increases in regional complexity can lead to significant advancements in proximity to more complex sectors. In a context where development is heavily influenced by path dependence, regional polarization tends to escalate rapidly, exacerbating existing disparities.

The econometric tests confirmed the hypothesis of the principle of relatedness, while indicating the influence of sectoral complexity on diversification, considering the differentiation of regions based on their complexity levels. The regions were grouped to examine the effect of relatedness density and PCI on the probability of entry or exit of sectors from local portfolios. Consistent with expectations, all estimated models demonstrated a positive relationship between density and the probability of entry, as well as a negative relationship between density and the probability of exit. However, the impact of PCI has proven to be crucial to understand. In highly complex regions, an increase in sectoral complexity raised the probability of new sector entry, while having minimal impact on the probability of sector exit. Conversely, in less complex regions, greater sector complexity was found to decrease the probability of entry and significantly contribute to sector exit.

In this regard, this paper addresses a significant gap in the existing literature on the subject. Prior studies primarily focused on evaluating hypotheses related to the principle of relatedness, while giving less attention to the influences of sectoral and regional complexity. Previously, there was no disaggregated differentiation of regions based on complexity, and consequently, there was a lack of a comprehensive analysis of the combined influence of relatedness density and PCI on the probability of sector entry and exit in the local portfolio. Therefore, the main contribution was to propose an assessment of regional diversification in Brazil with a focus on complexity, by introducing a more detailed differentiation of regions based on their complexity levels and examining the impact of sectoral complexity on the diversification patterns within each group.

The findings emphasize that the loss of complex activities poses a primary challenge for less complex regions. This dynamic restricts their diversification opportunities, increasing the gap regarding more complex regions. These findings suggest that diversification into more complex (unrelated) activities is very risky when it comes to low complexity regions. Moreover, the effect of entering activities with higher complexity becomes weaker in regions

at the medium-high complexity level, indicating that this is the least risky development level to pursue unrelated diversification, while medium-low complexity regions could be the target of hybrid strategies, since the exit risk reduces considerably in relation to low complexity regions.

However, the discussion does not stop here. There is still a limited amount of research in the literature that aims to evaluate relatedness from this perspective. While it has been established why related diversification is good news only for some regions, further questions arise to delve deeper into this analysis. How good or how bad is this news? What policies can effectively support the development of less complex regions? Furthermore, it is crucial to assess the results considering the methodological limitations of the research. The database used exclusively covers formal jobs, and the definition of the ECI complexity groups originated from arbitrary values. Future estimates should take into account the relevant informal scope in Brazilian labor market; further explore the categorization of regions based on complexity, and test sector entry and exit events adopting alternative criteria. Additionally, case studies would provide valuable insights into understanding the potential obstacles faced by low complexity regions and exploring the variations among high complexity regions that exhibit different levels of proximity to other complex sectors.

Notes

1. See more in [Hidalgo and Hausmann \(2009\)](#).
2. Microregions are geographic areas consisting of neighboring municipalities that share similarities in terms of spatial organization. These regions are characterized by specific features related to the agricultural, industrial, mineral extraction and fishing production structures, as established by [IBGE \(1990\)](#).
3. Furthermore, the option for presenting the logit model for the microregion groups is that the coefficients for PCI are slightly more significant.

References

- Balassa, B. (1965). Trade liberalization and revealed comparative advantage. *Manchester School of Economics and Social Studies*, 33(2), 99-123.
- Boschma, R., & Iammarino, S. (2009). Related variety, trade linkages, and regional growth in Italy. *Economic Geography*, 85(3), 289–311. doi: [10.1111/j.1944-8287.2009.01034.x](https://doi.org/10.1111/j.1944-8287.2009.01034.x).
- Boschma, R., Heimeriks, G., & Balland, P. A. (2014). Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy*, 43(1), 107–114. doi: [10.1016/j.respol.2013.07.009](https://doi.org/10.1016/j.respol.2013.07.009).
- 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), 223–250. doi: [10.1093/icc/dtu012](https://doi.org/10.1093/icc/dtu012).
- Essletzbichler, J. R. (2013). Industrial branching and technological cohesion in US metropolitan areas. *Regional Studies*, 49(5), 752–766. doi: [10.1080/00343404.2013.806793](https://doi.org/10.1080/00343404.2013.806793).
- Florence, S. P. (1948). *Investment, location and size of plant*. Cambridge: Cambridge University Press.
- Françoso, M. S., Boschma, R., & Vonortas, N. (2022). *Regional diversification in Brazil: The role of relatedness and complexity. Papers in evolutionary economic geography (PEEG)*. Utrecht: Utrecht University.
- Freitas, E. (2019). *Indústrias relacionadas, complexidade econômica e diversificação regional: Uma aplicação para microrregiões brasileiras*. Tese (Doutorado). Programa de Pós-Graduação da Faculdade de Economia da Universidade Federal de Minas Gerais.

-
- Freitas, E. E., & Paiva, E. A. (2016). Diversificação e sofisticação das exportações: Uma aplicação do product space aos dados do brasil. *Revista Econômica Do Nordeste*, 46(3), 79–98. doi: [10.61673/ren.2015.192](https://doi.org/10.61673/ren.2015.192).
- Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5), 685–697. doi: [10.1080/00343400601120296](https://doi.org/10.1080/00343400601120296).
- Glaeser, E., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100(6), 1126–1152. doi: [10.1086/261856](https://doi.org/10.1086/261856).
- Hartmann, D., & Pinheiro, F. (2022). Economic complexity and inequality at the national and regional level. In P. Chen, W. Elsner, A. Pyka, & Org (Eds.), *Handbook of Complexity Economics*. [S.l.: s.n].
- Hartmann, D., Bezerra, M., Lodolo, B., & Pinheiro, F. L. (2020). International trade, development traps, and the core-periphery structure of income inequality. *Economía*, 21(2), 1441–1452. doi: [10.1016/j.econ.2019.09.001](https://doi.org/10.1016/j.econ.2019.09.001).
- Hausmann, R., & Klinger, B. (2007). The structure of the product space and the evolution of comparative advantage. *CID Working Papers*, 146.
- Henderson, J. V., Kuncoro, A., & Turner, M. (1995). Industrial development in cities. *Journal of Political Economy*, 103(5), 1067–1085. doi: [10.1086/262013](https://doi.org/10.1086/262013).
- Hidalgo, C., & Hausmann, R. (2009). The building blocks of economic complexity. In *PNAS*.
- Hidalgo, C., Klinger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837), 482–487. doi: [10.1126/science.1144581](https://doi.org/10.1126/science.1144581).
- Hidalgo, C., et al. (2018). *The Principle of relatedness*. Berlin: Springer Proceedings in Complexity.
- IBGE (1990). Divisão regional do brasil em mesorregiões. *e microrregiões geográficas. v., 1*.
- Magalhaes, J. C. R., & Alves, P. J. H. (2021). A relação entre o crescimento econômico e as desigualdades regionais no brasil. *Texto para Discussão*, 1–46. 2621. doi: [10.38116/td2621](https://doi.org/10.38116/td2621).
- 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), 237–266. doi: [10.1111/j.1944-8287.2011.01121.x](https://doi.org/10.1111/j.1944-8287.2011.01121.x).
- Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge, MA, London: The Belknap Press.
- Neto, A. M. (2014). *Desigualdades regionais no brasil: Características e tendências recentes* (pp. 67–81). Brasília: Boletim regional, urbano e ambiental do IPEA.
- Pinheiro, F. et al. (2022). The dark side of the geography of innovation. relatedness, complexity and regional inequality in Europe. In *Papers in Evolutionary Economic Geography*, p. 1–47.
- Rezende, E. H., et al. (2023). Complexidade econômica e emprego formal no brasil: Evidências de regressão produtiva entre 2006 e 2020. In *Mercado de Trabalho: conjuntura e análise, Boletim IPEA*. Brasília (pp. 61–78). [s.n].
- Rigby, D. L. (2013). Technological relatedness and knowledge space: Entry and exit of us cities from patent classes. *Regional Studies*, 49(11), 1922–1937. doi: [10.1080/00343404.2013.854878](https://doi.org/10.1080/00343404.2013.854878).
- Romero, J., et al. (2022). Economic complexity and regional development: Evidence from brazilian municipalities. In *Proceedings of the EAEPE Conference* (pp. 1–25).
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 5(1), 379–423. doi: [10.1145/584091.584093](https://doi.org/10.1145/584091.584093).
- Silva, S. A. (2017). *da. Regional inequalities in Brazil: Divergent readings on their origin and public policy design*. EchoGeo.
- Vreyer, P. D., & Spielvogel, G. (2009). Spatial externalities between brazilian municipios and their neighbours. In M. PRESS (Ed.), *Poverty, Inequality, and Policy in Latin America*. [S.l.: s.n].

Annex

The supplementary material for this article can be found online.

Relatedness
and regional
economic
complexity

Corresponding author

Arthur Ribeiro Queiroz can be contacted at: queirozrthur@gmail.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com