Spatial Economic Analysis

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rsea20

Complexity-based diversification strategies: a new method for ranking promising activities for regional diversification

João P. Romero, Elton Freitas, Fabrício Silveira, Gustavo Britto, Fernanda Cimini & Frederico G. Jayme Jr.

To cite this article: João P. Romero, Elton Freitas, Fabrício Silveira, Gustavo Britto, Fernanda Cimini & Frederico G. Jayme Jr. (22 Aug 2024): Complexity-based diversification strategies: a new method for ranking promising activities for regional diversification, Spatial Economic Analysis, DOI: 10.1080/17421772.2024.2384744

To link to this article: https://doi.org/10.1080/17421772.2024.2384744

SPATIAL ECONOMIC ANALYSIS
https://doi.org/10.1080/17421772.2024.2384744

Complexity-based diversification strategies: a new method for ranking promising activities for regional diversification

João P. Romero [a], Elton Freitas [b], Fabrício Silveira [a], Gustavo Britto [a], Fernanda Cimini [a] and Frederico G. Jayme Jr. [a]

ABSTRACT

Following the increasing upsurge in works using economic complexity to devise smart diversification strategies, this paper proposes a new complexity-based method to be used by policymakers to rank promising activities for short, medium and long-term diversification. After reporting the positive impact of regional complexity on employment and gross domestic product (GDP) per capita growth for Brazilian regions, the paper assesses the potential of the smart diversification score (SDS) to increase regional complexity. Looking backwards, the paper finds that SDS can predict up to 39.4% of the diversification of 1033 Brazilian cities that have increased complexity between 2007 and 2018. Looking forward, the paper calculates the SDS for the city of Belo Horizonte, suggesting a portfolio of related and unrelated activities, finding high potential gains to be obtained following alternative paths.

KEYWORDS

Regional development, smart diversification, economic complexity, economic growth

JEL O11, O47, R11

HISTORY Received 15 July 2023; in revised form 30 April 2024

1. INTRODUCTION

Economic development is intrinsically related to structural change. The classical literature on economic development highlights the importance of increasing the share of manufacturing in the economy, while reducing the share of agriculture (e.g., Furtado, 1964; Hirschman, 1958; Kaldor, 1966; Prebisch, 1962; Schumpeter, 1934). Structural change, therefore, involves learning and mastering new economic activities. As technology evolved, some sectors became more science-based than others and markets for manufacturing products changed markedly. Accordingly, modern approaches to economic development started stressing the importance of moving into high-tech manufacturing for sustainable development (e.g., Lundvall, 1992; Nelson, 1993; Romero & Britto, 2017).

With the rapid changes in science and technology observed in the past few decades, high-tech manufacturing industries have become increasingly more heterogeneous, requiring highly specialised knowledge. The path-breaking work of Hidalgo et al. (2007) explored fine-grained international trade data to build a network that interconnects products according to the

CONTACT João P. Romero jpromero@cedeplar.ufmg.br
[a] Center for Regional Development and Planning (Cedeplar), Department of Economics, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil
[b] Department of Economics, Universidade Federal de Sergipe (UFS), Aracaju, Brazil

Supplemental data for this article can be accessed online at https://doi.org/10.1080/17421772.2024.2384744

probability of competitive co-production. This network, the *Product Space*, indicates the proximity of the productive knowledge required to produce each pair of goods. Moreover, it makes it clear that development is heavily path-dependent due to the differences in accumulated knowledge between economies. As Britto et al. (2019) have shown, the shape of the *Product Space* has gradually changed over time, with clusters of products becoming clearer and more separated. This illustrates the increase in specialised knowledge that led to higher separations between manufacturing fields.

Exploring even further the information contained in disaggregate international trade data, Hidalgo and Hausmann (2009) showed that the ubiquity of the competitive production of different goods varies markedly. Furthermore, they also showed that the level of diversification of each economy is associated with its level of income per capita. Combining these two raw measures, the authors created the product complexity index (*PCI*) and the economic complexity index (*ECI*). The former indicates the amount of productive knowledge required to produce each good competitively. The latter indicates the amount of productive knowledge available in each economy. Hausmann, Hidalgo, et al. (2014) provided evidence that indicates that increasing economic complexity predicts considerably higher growth rates of income per capita in the future, even after controlling for several additional variables.

In parallel to the development of the economic complexity literature, several studies have also investigated the importance of diversification for regional development. Glaeser et al. (1992) have provided seminal evidence of the importance of regional diversification vis-à-vis specialisation for employment growth. A few years later, Frenken et al. (2007) showed that it is diversification into related activities that contributes to employment growth, while diversification into unrelated activities contributes to increased regional resilience, reducing unemployment growth.

In the last few years, therefore, some works have been combining the evidence on regional diversification with the evidence of the importance of economic complexity for regional development. At the regional level, patent and employment data have been used to compute economic complexity indicators. Balland et al. (2019), for example, use patent data to measure local technological knowledge, using the same methods proposed by Hidalgo et al. (2007) and Hausmann, Hidalgo, et al. (2014). The authors found that related diversification increases the probability of becoming competitive in a given industry. Moreover, the authors used this finding to guide the formulation of regional smart specialisation strategies.

Despite the mounting evidence on the importance of related diversification for regional growth, to the best of our knowledge, no study has yet investigated whether economic complexity is associated with gross domestic product (GDP) per capita and employment growth at the city level in developing countries. In the specific case of Brazil, Geleti et al. (2021) have investigated the importance of economic complexity for regional diversification, while Teixeira et al. (2022) have tested the impact of economic complexity, calculated using export data, on GDP growth, but only at the state level. Queiroz et al. (2023) have tested the effect of economic complexity on employment growth for Brazilian states, while Chávez et al. (2017) have used employment data to test the relationship between economic complexity and GDP per capita growth for Mexican states. Furthermore, even for developed countries the evidence is still scarce (Davies & Maré, 2021; Li & Rigby, 2023).

The reasoning for this analysis is straightforward. Although related diversification is found to be important for regional growth, diversification into more complex activities is crucial, as found in the literature on economic complexity at the country level. Furthermore, the high correlation between measures of industry cohesion, such as regional density (Hausmann, Hidalgo, et al., 2014) and economic complexity, provide initial evidence that it might be relevant for regional development.

The realised potential of the economic complexity approach has been shown by Hidalgo (2021), who stressed that the literature has branched out considerably towards many different areas and issues. In one branch of this literature, recent studies have been using indicators based on the economic complexity methodology to guide the design of development policies. Hausmann and Chauvin (2015) used a series of indicators constructed based on economic complexity and relatedness between products to identify promising sectors for the development of Rwanda, while Hausmann, Hidalgo, et al. (2014) used a similar methodology to identify diversification possibilities for Uganda. Soon after the research moved to the analysis of the diversification possibilities of regions in Panama (Hausmann et al., 2017), Brazil (Romero and Silveira, 2019; Queiroz et al., 2023), Mexico (Hausmann et al., 2020a) and Argentina (Hausmann et al., 2020b).

Nonetheless, there is still considerable room for improvement in the methods used to identify promising activities for economic diversification, as stressed by Hidalgo (2023). Most of the studies have sought to combine several variables to identify promising activities attributing ad hoc weights to each of them. Yet, using a better method to establish the weights of each variable would most likely improve the quality of the identified activities.

This paper provides three contributions to the literature on economic complexity and regional development using data from Brazilian municipalities and microregions. First, it proposes a new method to rank promising activities to be targeted by regional development policies, providing two improvements to previous works: (i) the method combines a series of indicators, as proposed by Hausmann et al. (2017), but uses weights estimated using principal component analysis and (ii) it identifies promising activities for short, medium and long-term diversification proposing a balanced portfolio between related and unrelated activities. This methodology is illustrated using the example of the Brazilian city of Belo Horizonte. A second contribution of the paper is to report econometric regressions of the impact of regional economic complexity, estimated using employment data from Brazil, on the growth rates of GDP per capita and of formal employment per capita. These regressions investigate whether the results found by Hausmann, Hidalgo, et al. (2014) are also valid at the regional level. To the best of our knowledge, this is the first paper to perform such analysis for Brazil at the most granular spatial units. Third, using the estimates of the relation between economic complexity, income and employment, it presents projections of the potential gains to be obtained following different development paths.

The remainder of the paper is organised as follows. Section 2 presents a literature review on economic complexity and regional development. Section 3 presents the methodology for devising complexity-base smart diversification strategies. Section 4 presents the empirical investigation on the relationship between economic complexity, GDP per capita and employment growth. Section 5 discusses the application of the methodology for the case of the city of Belo Horizonte. Section 6 present the concluding remarks.

2. REGIONAL ECONOMIC COMPLEXITY AND REGIONAL DEVELOPMENT

2.1. Economic complexity

The seminal paper by Hidalgo and Hausmann (2009) proposed to calculate the complexity of products and countries based on information on the diversification of economies and the ubiquity of products. The level of diversification of each country is defined as the number of products the country produces with revealed comparative advantage (*RCA*) above one, while the level of ubiquity of each good is defined as the number of countries that produce the good with *RCA* above one. Formally:

$$RCA_{cp} = \frac{x_{cp} / \sum_p x_{cp}}{\sum_c x_{cp} / \sum_c \sum_p x_{cp}} \tag{1}$$

$$Diversification = k_{c,0} = \sum_p M_{cp} \tag{2}$$

$$Ubiquity = k_{p,0} = \sum_c M_{cp} \tag{3}$$

where x denotes the export quantum, while subscripts c and p denote country and product, respectively. M is a *dummy* variable which equals one if country c exports the good p with RCA, and zero otherwise. An RCA above one shows that the country is competitive in the production of the good, while the opposite holds if the index is below one.

Hidalgo and Hausmann (2009) showed that there is a strong positive association between diversification and income per capita. In addition, they showed that diversification and ubiquity are negatively correlated, which suggests that countries that are more diversified tend to produce goods that are less ubiquitous.

Based on the information from these indexes, Hidalgo and Hausmann (2009) calculate a product complexity index (*PCI*) and an economic complexity index (*ECI*). A country with a high level of diversification is considered less complex if the products produced competitively (with *RCA*) present high ubiquity. Analogously, a product with a low level of ubiquity is considered more complex if it is produced by countries with high diversification. Consequently, by repeating this process and performing continuous iterations between the two indexes it is possible to extract more refined information about the economic complexity of each country and product. Formally:

$$k_{c,N} = (1/k_{c,0}) \sum_p M_{cp} k_{p,N-1} \tag{4}$$

$$k_{p,N} = (1/k_{p,0}) \sum_c M_{cp} k_{c,N-1} \tag{5}$$

where N denotes the number of iterations.

Substituting (4) into (5) yields:

$$k_{c,N} = \sum_{c'} \tilde{M}_{cc'} k_{c',N-2} \tag{6}$$

where $\tilde{M}_{cc'} = \sum_p (M_{cp} M_{c'p})/(k_{c,0} k_{p,0})$ and c' denote other countries besides c.

Equation (6) is satisfied when $k_{c,N} = k_{c,N-2} = 1$. This is the eigenvector associated with the highest eigenvalue of $\tilde{M}_{cc'}$, which is formed of ones and, therefore, is uninformative. Consequently, the eigenvector associated the second highest eigenvalue ($\vec{K}$) of $\tilde{M}_{cc'}$ is used to capture the highest part of the variance. Hence, *ECI* is calculated as:

$$ECI = \left(\vec{K} - \left\langle \vec{K} \right\rangle\right) / sd(\vec{K}) \tag{7}$$

where < > denotes the average and *sd* denotes the standard deviation.

An analogous procedure is used to calculate *PCI*, but substituting (5) in (4) and using the eigenvector associated with the second highest eigenvalue ($\vec{Q}$) of the matrix $\tilde{M}_{pp'}$:

$$PCI = \left(\vec{Q} - \left\langle \vec{Q} \right\rangle\right) / sd(\vec{Q}) \tag{8}$$

Following their seminal results, a series of papers explored other effects of economic complexity. Hausmann, Hidalgo, et al. (2014) provided evidence that indicates that increasing economic complexity predicts considerably higher growth rates and levels of GDP per capita in the future, even after controlling for a series of additional variables. Hartmann et al. (2017), found evidence that economic complexity is correlated with lower income inequality. Mealy and Teytelboym (2021) and Romero and Gramkow (2021) found strong evidence indicating that increasing economic complexity contributes to reduce greenhouse gas emissions and other environmental impacts.

2.2. Product space

The pathbreaking paper of Hidalgo et al. (2007) explored whether the sectoral composition of each country's competitive exports influences the path of its structural transformations. In this paper the authors explore the idea that each country's productive structure influences its growth and development possibilities, stressing the path-dependence of knowledge and capabilities accumulation. Hidalgo et al. (2007) assume that the probability of producing two products that require similar capabilities for their competitive production is higher than the probability of producing two goods that require different capabilities. The proximity between two products (p and j) is given by:

$$\varphi_{p,j} = min\{P(RCA_j = 1|RCA_p = 1), P(RCA_p = 1|RCA_j = 1)\} \tag{9}$$

Adopting a threshold value for proximity, the authors established the active links between products, creating a network that they called *Product Space*. In this network, products that use similar capabilities for their competitive production tend to form clusters. Moreover, complex products tend to be located towards the centre of the network, while least complex products tend to locate towards the fringes.

Using the *Product Space*, Hidalgo et al. (2007) showed that, on average, less developed countries produce goods with a limited number of links. This restricts these countries' diversification possibilities, making it more costly for these countries to move towards the production of more complex products. The opposite holds true for developed ones. Hence, they show that different countries face different opportunities to diversify their economies and increase their economic growth, given their distinct productive structures and associated capabilities. Moreover, these results indicate that achieving competitiveness in the production of complex goods takes time, since it requires learning new capabilities (Hidalgo et al., 2007, p. 487).

Seeking to explore the implicit information contained in the *Product Space*, Hausmann, Hidalgo, et al. (2014) developed an indicator that measures the ease of acquiring competitiveness in any given industry as a function of existing capacities in the economy. This indicator would help to identify new diversification possibilities based on the implied costs associated with the acquisition of the new capabilities required for performing new activities competitively. This index, called product density index (*PDI*), measures the proximity of a given product to the country's current production structure (products with *RCA*), indicating the difficulty (or cost) of achieving *RCA* in this product. Hence, this measure reflects the amount of new productive knowledge that a region needs to acquire to be able to produce and export a particular product with *RCA*. Formally:

$$PDI_{pct} = \frac{\sum_p M_{ict}\ \varphi_{pi}}{\sum_p \varphi_{pi}} \tag{10}$$

Hausmann, Hidalgo, et al. (2014) also proposed a second indicator, called the opportunity gain index (*OGI*). This index measures the gains that achieving *RCA* in any given good brings in terms of facilitating the production of more complex goods that were not previously produced/exported competitively in this economy. Formally:

$$OGI_{pct} = \sum_{p} \frac{(1 - M_{cit})\varphi_{pi} PCI_{it}}{\sum_{p} \varphi_{pi}} - (1 - PDI_{pct}) PCI_{pt} \quad (11)$$

A high value in the OGI, therefore, indicates that the product under investigation is closer to complex products. Hence, this index can be used to devise policies that aim to increase an economy's economic complexity by considering several rounds of diversification into progressively more complex goods.

2.3. Activity space

At the regional level using export data to measure economic complexity is highly problematic since transactions between regions within the same country are not computed. Furthermore, economic interactions between neighbours are stronger at the regional level, making knowledge spillovers more relevant. In addition, in regions services play a more prominent role, so that it becomes more relevant to take these activities into account.

To solve these issues, some studies have been using employment or patent data instead of trade data to calculate the indicators of economic complexity. Using employment data has one additional advantage: it makes it possible to use information on the number of occupations within companies or regions to measure proximity through occupational similarities (Farjoun, 1994, p. 188). Moreover, a growing number of works have been using occupations and skill to measure regional complexity, finding evidence of its relevance for regional diversification and growth (e.g., Buyukyazici et al., 2023; Lo Turco & Maggioni, 2022).

Following Freitas (2019), from the concept of co-occupation it is possible to estimate the proximity of activities with similar jobs and build the complexity indicators through employment data. First, we define the indicator of effective occupations (EO), analogous to the revealed comparative advantage index (RCA), as the basis for calculating the complexity indicators using employment data. Formally:

$$EO_{s,o} = \frac{emp_{s,o}/emp_{s}}{emp_{o}/emp} \quad (12)$$

where $emp_{s,o}$ is the employment of occupation o in sector s and emp_s is the total employment of sector s in the country. In addition, emp_o is the total employment of the occupation o in the country and emp is the total employment in the country.

Thus, if the EO indicator is equal or greater than one, the contribution of occupation o in sector s is greater than the participation of occupation o in the country. Hence, the sector in question effectively employs such an occupation. Otherwise, if EO is less than one, the conclusion is that the sector does not effectively employ this occupation in the analysed location.

Using the EO indicator it is possible to estimate the proximity between activities and for an *Activity Space*. Proximity is calculated as the probability of an industry employing a certain occupation, given that another industry also employs that occupation. This represents a different way, although similar, of measuring the similarities between industries in terms of occupations. Thus, following Freitas (2019), Equation (9) can be adapted to establish the relationship between activities s and i replacing RCA by EO.

The strategy of using co-occurrence of occupations in each industry was employed to establish the links between industries because this method resulted in a more coherent description of the connections between sectors than the links established though co-occurrence of industries with RCA in each region.

2.4. Regional development

Understanding how new regional growth paths emerge has been repeatedly raised by economic geographers as one of the most intriguing and challenging questions in the field (Martin & Sunley, 2006; Scott, 1988; Storper & Walker, 1989). After all, the industrial history of regions is expected to affect how regional structures give rise to new activities over time, and how they transform and restructure their economies.

The issue of structural change in regional and urban economics gained new impetus through recent reassessments of the works of Jane Jacobs and Alfred Marshall. Glaeser et al. (1992) provided important contributions to the framework of agglomeration economies, by presenting evidence of the economic importance of the diversification of urban areas. Henderson et al. (1995) found that new industries, especially the high-tech, entered diversified cities where Jacobs' externalities were available, while mature industries benefited more from location externalities generated in more specialised cities. According to Simões and Freitas (2014), Jacobs' externalities are more relevant in sectors of high technological intensity, while sectors with low and medium technological intensity benefit more from locating in medium-sized urban centres, relatively less diversified.

Jacobs' main argument for new industries benefiting from diversified urban economies was that urban diversity prompts the division of labour in the city. However, the division of labour contributes to urban growth due to its effect on opportunities for innovation, and not so much due to its effect on technical efficiency. This fits very well into the Schumpeterian concept of innovation as successful new combinations of productive forces, i.e., old ideas.

Cognitive theory, however, emphasises a trade-off between diversity and similarity. Although cognitive proximity (overlapping competencies) facilitates communication between agents, only the ones who do not share overlapping competencies and knowledge can offer something new to be learned by its counterpart (Nooteboom, 2000). Hence, the fact that social learning may require an optimal level of cognitive distance may explain why, after several empirical studies, the evidence on the effects of Jacobs externalities is still inconclusive (De Groot et al., 2009).

Regional knowledge spillovers only happen between certain industries, since more effective communication is often hampered by cognitive distance. Recently, several authors have suggested that industries are more likely to learn from each other when they are technologically related (Almeida & Kogut, 1999; Boschma & Frenken, 2009, 2011; Gilsing et al., 2007). Thus, a diversified portfolio of technologically related industries should be more beneficial for regional growth than a diversified portfolio of industries in a broad range of technological areas, given that it is the combination of distance and cognitive proximity that brings together the positive sides of diversity and similarity across industries.

Frenken et al. (2007) argue that regions with a greater degree of variety in related industries present more learning opportunities and, consequently, higher dissemination of local knowledge. Using data for the Dutch economy, the authors show that regions with a higher degree of 'related variety' present higher employment growth. The same result was also found for other countries (Bishop & Gripaios, 2009; Essletzbichler, 2005). Moreover, Boschma and Iammarino (2009) argue that related variety can also flow from one region to another through commercial links between industries. Using regional trade data, the authors show that inter-regional knowledge flows are associated with regional employment growth, when these come from industries related to industries in the region.

According to Boschma and Frenken (2009), the process through which new activities arise from technologically related industries can be termed 'regional branching'. The reason this process of regional branching takes place is that new industries can connect to existing ones through various knowledge transfer mechanisms. These mechanisms are: (i) diversification of firms; (ii) entrepreneurship in the form of spinoffs; (iii) mobility of workers and (iv) social networks.

The branching process is essentially a regional phenomenon, as these mechanisms operate primarily, although not exclusively, at the regional level, that is, within subnational regions rather than across regions.

Nonetheless, despite the mounting evidence on the importance of related diversification for regional growth, the number of papers that have investigated the importance of economic complexity for regional development is relatively lower than the ones on technology complexity or related diversification (Balland et al., 2022). At the regional level, patent and employment data have been used to compute economic complexity indicators, since exports are not as informative as they are at the national level because trade between regions within the same country is not computed.

Balland et al. (2019), for example, use patent data to measure local technological knowledge, employing the same methods proposed by Hidalgo et al. (2007) and Hausmann, Hidalgo, et al. (2014). Using data from the European Union, the authors found that higher knowledge complexity and related diversification increase the probability of becoming competitive in a given industry. Moreover, the authors used these findings to guide the formulation of regional smart specialisation strategies. Nonetheless, they do not test the impact of technological complexity on regional GDP per capita growth. This investigation was carried out by Li and Rigby (2023) who, using patent data, provided evidence that the technological complexity of Chinese regions had a positive and significant impact on GDP growth.

Using employment data from Mexican states, Chávez et al. (2017) found that economic complexity predicted higher GDP per capita growth. In a similar analysis using employment data from New Zealand, Davies and Maré (2021) found that economic complexity and relatedness impacted on employment growth, but only in large cities.

In the case of Brazil, specifically, Geleti et al. (2021) used employment data from Brazil to calculate the skill relatedness between activities. They tested the relationship between skill relatedness, entry, exit and employment growth in a panel of microregions and found that relatedness is in general positively associated with entry and employment growth, and negatively associated with exit. Moreover, they also found a similar association between economic complexity, entry, exit and employment growth. As in the case of Balland et al. (2019), Geleti et al. (2021) did not test the impact of economic complexity on regional GDP per capita growth either. Using export data from Brazilian states, Teixeira et al. (2022) found that economic complexity had a positive and significant impact on GDP growth, while Queiroz et al. (2023) found that economic complexity predicted higher employment growth. Both studies, however, used a high level of aggregation for economic activities (CNAE at two digits). Finally, Gao et al. (2021) apply the economic complexity methodology using employment data from both China and Brazil, to show that knowledge spillovers are relevant at the regional level, and that improving transport infrastructure helps increasing these spillovers and the productive diversification they foster.

Despite this evidence, however, no paper has yet explored whether economic complexity is associated with GDP per capita and employment growth at lower regional levels (municipality and microregion) in Brazil. The reasoning for this investigation is straightforward. Although related diversification is found to be important for regional employment growth, diversification into more complex activities might also be relevant for regional development, as found in the literature of economic complexity at the country level. Furthermore, the high correlation between measures of industry cohesion such as regional density and economic complexity provide initial evidence that economic complexity might be relevant for regional development.

Finally, several works have used economic complexity and relatedness indicators to identify promising activities for regional development. The first of such works was Hausmann, Hidalgo, et al. (2014), which used a series of indicators constructed based on economic complexity and on relatedness between products to identify promising sectors for the development of Uganda. Hausmann and Chauvin (2015) and Romero and Freitas (2018) used a similar method to identify promising sectors for the development of Rwanda and Brazil, respectively. Hartmann et al. (2019), on the other hand, explored different rules to establish diversification strategies for Paraguay, taking into account not only their growth prospect but also their impact on reducing inequality. Moreover, similar approaches have also been applied to the analysis of the diversification possibilities of regions in Panama (Hausmann et al., 2017), Brazil (Romero and Silveira, 2019; Queiroz et al., 2023), Mexico (Hausmann et al., 2020a) and Argentina (Hausmann et al., 2020b).

Moreover, economic complexity indicators have also been used to analyse and propose smart specialisation policies. Balland et al. (2019), for instance, proposed a framework for the European smart specialisation policy based on relatedness and complexity indexes calculated using patent data. Buyukyazici et al. (2023), on the other hand, used measures of skill relatedness and skill complexity to assess the European smart specialisation policy, and found that regions have been focusing more on skill relatedness than on skill complexity.

As stressed by Pinheiro et al. (2022a; 2022b), however, there is a downside in related diversification. Countries and regions at low levels of economic complexity tend to diversify into low-complexity activities while the opposite holds true for countries and regions at high levels of complexity. This tends to increase income disparities. Hence, it becomes crucial for underdeveloped regions to implement unrelated diversification in order to improve their development paths. Yet, as Pinheiro et al. (2022b) has shown, unrelated diversification becomes more relevant for countries with intermediate complexity levels. This evidence is directly connected to the middle-income trap literature. Hartmann et al. (2021) show that while some countries managed to perform this transition, most were not able to do so. Hence, as stressed by Hidalgo (2023), development policies should focus both on related and unrelated diversification, although at different proportions in different moments of the development trajectory of the country.

The main contribution of this paper to this literature is to propose a method to improve the selection of promising sectors for regional diversification in two ways: (i) by using principal component analysis to weight the variables considered and (ii) by dividing promising activities into different subgroups in order to promote both related and unrelated diversification, organised into short, medium and long-term strategies.

3. COMPLEXITY-BASED DIVERSIFICATION STRATEGIES

3.1. Database

The empirical investigation presented in this paper is based on employment data for the municipalities and microregions of Brazil over the period 2006–2017. Employment data used for the construction of the *Activities Space* and of the economic complexity indicators of local activities come from RAIS (Annual Report of Social Information), of the Brazilian Ministry of Employment. The database provides information on the number of employees by economic activity classified according to the CNAE (version 2.0) classification and by occupation within each activity at both regional levels explored here. The CNAE classification comprises 672 activity classes from all sectors (services, manufacturing and agriculture). Moreover, employment is divided into 596 occupations within each activity. The database covers the 5570 Brazilian municipalities, which are grouped into 558 microregions, which represent the effective economic areas according to IBGE (Brazilian Institute of Geography and Statistics).

Brazilian foreign trade data were gathered from SECEX and UN COMTRADE classified by product according to the Mercosur Common Nomenclature (NCM) classification. Import data was used as a proxy of the demand for the output of each activity. A correspondence table provided by IBGE was used to associate NCM products to economic activities classified by CNAE (version 2.0). Four-year averages of the variables were calculated to smooth short-term fluctuations and reduce eventual measurement errors. Year 2006 was dropped from the analysis.

3.2. The Brazilian activity space

Using the employment information from the RAIS database it is possible to build the *Activity Space* for Brazil following the methodology presented in Section 2.3. The network,[1] displayed in Figure 1, presents the active links between the activities from the CNAE classification, so that each node represents an activity. Figure 1 shows that manufacturing activities (in brown) and modern services (in dark green) are more connected and localised towards the centre of the network, highlighting their importance for diversification and development.

3.3. Smart diversification score

Several studies have sought to use the indicators discussed in the previous section to devise diversification policies (e.g., Hausmann et al., 2017; Hausmann & Chauvin, 2015; Romero & Freitas, 2018). These studies gathered a series of variables related to three relevant dimensions related to each product not exported with *RCA*: (i) *Markets*; (ii) *Capabilities* and (iii) *Gains* (in terms of economic complexity). The variables were then normalised and put together in a weighted sum to generate a score of promising industries to be fostered through development policies. The main problem with this methodology is that the weights attributed to each variable are completely arbitrary. Hence, in this paper, we seek to solve this problem and improve the economic complexity smart diversification score by using principal component analysis (PCA).

PCA is used to explain the structure of variance and covariance of a random vector, composed of p-random variables, through the construction of a linear combinations of the original variables. It allows the researcher to reorient the data so that the first few dimensions explain as much information as possible. The individual main components are linear combinations of random variables X_1, X_2, Xn. Geometrically, these linear combinations represent the selection of a new coordinate system obtained by rotating the original system with X_1, X_2, Xn as the coordinate axes. Hence, PCAs are used to discover and interpret dependencies that exist between variables and to examine relationships that may exist between individuals.

The procedure carried out to estimate the weights of the variables followed a series of steps. First, sub-samples of municipalities/microregions were defined based on the characteristics of their productive structures in terms of high/low density and complexity. Second, the weights

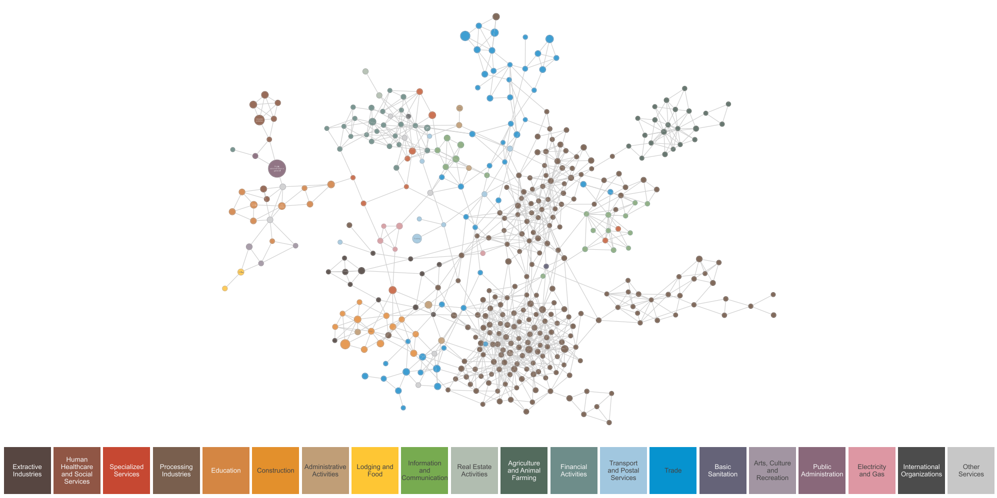

Figure 1. Activity space for Brazil
Source: Authors own elaboration based on employment data from RAIS.

attributed to each variable within each dimension (*Capabilities*, *Markets* and *Gains*) were defined by applying the PCA method. It was decided to keep the 'n' components generated that cumulatively explained at least 80% of the variance of each one of the dimensions.

Table 1 shows the variables considered to calculate the smart diversification score (*SDS*) and the weights estimated using PCA, which was carried out considering the sample of municipalities/microregions belonging to the high-density (average $PDI >$ median PDI) and high-complexity ($ECI > 0$) quadrant and which presented growth in the *ECI* in the three periods evaluated in this research. In other words, only the information of a relevant subgroup of regions was used to calculate the weights. In this case, considering that the score would be applied to a region in the high-density and high-complexity group. To calculate the weights, the variables referring to the most recent period available in the database were used, from 2015 to 2018.

To assess the predictive capacity of the *SDS* proposed in this paper, we investigated the capacity of the rule to predict the activities that acquired *RCA* in municipalities that presented an increase in their economic complexity during the period under analysis. First, we created rankings for the activities identified as promising by each of the *SDS* in period 1. Then, we selected the industries in which the municipalities that have increased their *ECI* did not have *RCA* in period 1 (2007 to 2010) and achieved *RCA* in period 3 (2015 to 2018). Finally, from the number of activities found in the previous step, we verified how many of them were on the top of the ranking built using the *SDS*. The limit was defined by the number of activities that transitioned from $RCA < 1$ to $RCA > 1$. For example, if a municipality X had 14 activities that had reached $RCA > 1$ in period 3, we evaluated the top 14 activities indicated by the *SDS* to identify how many of the activities were correctly identified as promising. Similarly, if a municipality Y had 30 activities that reached $RCA > 1$ in period 3, we verified

Table 1. Variables considered in the smart diversification score (SDS).

Dimensions	Weights	Indicators	Weights via PCA
Capabilities	0.33	Employment	0.359
		Revealed comparative advantage (*RCA*)	0.583
		Employment growth rate	0.058
Markets	0.33	Imports (municipality/microregion)	0.104
		Revealed comparative disadvantage (*RCD*) (municipality/ microregion)	0.040
		Import growth (municipality/microregion)	0.042
		Imports (state)	0.120
		Revealed comparative disadvantage (*RCD*) (state)	0.093
		Import growth (state)	0.216
		Imports (country)	0.132
		Revealed comparative disadvantage (*RCD*) (country)	0.102
		Import growth (country)	0.150
Gains	0.33	Activity density index (*ADI*)	0.216
		Activity complexity index (*ACI*)	0.293
		Opportunity gain index (*OGI*)	0.271
		Number of connections in the network	0.220

Source: Authors own elaboration.

the top 30 activities indicated by the *SDS*. From these comparisons we calculated the average rate of success the score as the ratio of the number of activities found in the final step of the process in relation to the total number of new activities with *RCA* for each municipality.

Table 2 shows the mean percentages of correct predictions for the *SDS* in comparison to the activities identified as promising using equal weights for each variable (within each dimension) instead of the weights calculated using PCA. Column 1 shows the percentages of the two rules if applied to all municipalities. Both rules have a similar rate of correct predictions, around 30%, which indicates that about one in every three activities that actually obtained *RCA* in period 3 were targeted as promising by the rules. Column 2 presents the percentages only for the sectors that gained *RCA* in period 3 and that had $PCI > 0$, that is, above average. As can be seen, in this case the results improve considerably: the *SDS* predicts 39.4% of the activities, 1.4 percentage points better than the rule that applies equal weights to all variables. The same pattern is observed in columns 3 and 4. The difference is that now the rates are assessed considering only the municipalities with high density and high complexity, which was the subsample used to calculate weights using PCA. The results are similar to the ones found in columns 1 and 2, with similar improvements found.

The results suggest the importance of the refinements carried out in the research, indicating that the use of weights calculated via PCA increase the proportion of sectors predicted as promising by the scores and that effectively achieved RCA in municipalities where an increase in economic complexity was observed.

It is important to note, however, that the assessment presented in Table 2 does not indicate the quality of the proposed smart diversification strategy (SDS). It is very likely that the percentages reported in Table 2 are relatively low because the actual diversification path followed by the municipalities that have increased their complexity is in fact sub-optimal in relation to the diversification path proposed by the *SDS*. In other words, the percentages might be low not because of a problem of the *SDS*, but because of its quality. This is made clear by the comparison of columns 1 and 3 with columns 2 and 4. The percentage of correct predictions of the *SDS* increases with the introduction of a filter to focus on activities with a product complexity index (PCI) above average. This shows that without this filter the percentage is lower because there has been in fact more diversification towards activities with below average complexity than what would be desirable.

Table 2. Assessment of the smart diversification score (SDS) for municipalities with increasing economic complexity index (ECI).

	Test 1	**Test 2**	**Test 3**	**Test 4**
	All municipalities	**All municipalities & Products with $PCI > 0$**	**All municipalities with high density and high ECI**	**All municipalities with high density and high ECI & Products with $PCI > 0$**
Ranking 1 (equal weights)	31.6%	38.0%	30.6%	35.7%
Ranking 2 (PCA weights)	30.5%	39.4%	30.0%	37.0%
Number of observations	4440	1033	635	362

Source: Authors own elaboration.

3.4. Alternative strategies

To improve even further the smart diversification strategies established following the score proposed in the previous sub-section, activities were classified according to four strategies, following a classification proposed by the Brazilian National Export Plan (2015–2018), namely: (i) *Recovery*, (ii) *Consolidation* and (iii) *Strategic Bets*. The scores were used to rank the most promising sectors within each of these strategies. The purpose of this ranking is to bridge the gap between the proposed methodology for the selection of promising sectors and different smart diversification strategies, based on extra information about characteristics of local competitiveness and of sectoral market dynamics.

Table 3 shows the parameters of the proposed typology and strategies. Through the typology presented in Table 3 it becomes possible to outline short, medium and long-term development strategies. Notice that a fourth type was included: *Maintenance*. The latter is not considered a development strategy since these are well developed in the municipality/microregion. This subgroup is introduced to stress the importance of keeping the competitive position of the region in these key activities.

It is important to stress that the criteria listed in Table 3 represent a second layer, applied after determining the promising activities using the smart diversification score (SDS) presented in Table 1. Short-term strategies should focus on strengthening activities classified as *Recovery*, since it encompasses sectors that were once competitive in the municipality but are now declining in the region. The *Consolidation* strategy is associated with medium-term development strategies, since it focuses on sectors that the municipality already has a certain level of competitiveness in but does not yet have *RCA*. Finally, the *Strategic Bets* strategy is associated with long-term strategies, as it considers promising high-complexity sectors in which the region's competitiveness is still low.

It is crucial to note that the different strategies aim to create a portfolio of activities that take both related and unrelated activities. While the short and medium-term activities associated with the *Recovery* and *Consolidation* strategies promote related diversification, the long-term activities associated with the *Strategic Bets* strategy are more focused on unrelated diversification. Although this is not explicit in the rules presented in Table 3, it is well known that for regions that do not present high economic complexity, activities with lower *RCA*, tend to be the ones with higher complexity. This information, along with the others, is nonetheless considered through the ranking constructed using the smart diversification score.

Table 3. Smart diversification strategies – subgroups.

Subgroups	Definition	Parameters
Maintenance	Sectors that are well positioned in the market and have a comfortable situation in relation to their main competitors	$RCA >= 1.5$ and municipal and national employment growth > 0
Recovery (Short-term)	Sectors that have not yet consolidated or products that were once consolidated but have been reducing their market share.	$RCA >= 0.5$ and employment growth < 0 and national > 0
Consolidation (Medium-term)	Sectors that are not yet consolidated but are growing at a pace close to or above that of their competitors.	$0.5 =< RCA < 1.5$ and Municipal and national employment growth > 0
Strategic Bets (Long-term)	Sectors whose participation is very low, but whose exports are growing in the market.	$0 < RCA < 0.5$ and municipal and national employment growth > 0

Note: Based on the National Export Plan 2015–2018. Activities that did not fit into subgroups were discarded.
Source: Authors own elaboration.

4. EMPIRICAL INVESTIGATION

4.1. Econometric specification

To analyse the impact of the *ECI* on economic growth at the regional level, we estimate a set of regressions in which the dependent variable is either the annualised growth rate of GDP per capita, or the growth rate of the share of formal employment in total population. *ECI* was calculated following the methodology presented in Section 2.1 but using employment data instead of export data to calculate RCA in each activity (CNAE) in each municipality or microregion.

The estimated equation for GDP per capita follows the specification used by Hausmann, Hidalgo, et al. (2014), but in a panel data form:

$$\log(y)_{i,t} = \alpha + f_i + f_t + \beta_1 ECI_{i,t-1} + \beta_2 ECI_{i,t-1} * \log(y)_{i,t-1} + \beta_3 \log(y)_{i,t-1} + \beta_i \log(x)_{i,t-1} + \varepsilon \tag{13}$$

where y is the GDP per capita (gathered from the Brazilian Institute of Geography and Statistics – IBGE), f are fixed effects for individuals i (municipalities/microregion) and periods t (annual dummies), α is the constant, ε is the residuals. Among the explanatory variables are the initial GDP per capita (log) and a multiplicative variable between *ECI* and GDP per capita. The former seeks to capture the effect of the hypothesis of convergence or technological catch-up. The multiplicative term seeks to capture the non-linearity of the effect of *ECI* on the growth rate. Hypothetically, this effect is negative because the potential gains from increasing *ECI* also reduce with the increase in GDP per capita and *ECI* over time. Finally, x indicates three control variables introduced in the regressions: (i) population and (ii) the percentage of workers with tertiary education.

The explanatory variables are introduced with a lag to assess if they predict increases in GDP per capita in the subsequent period. Although this does not solve potential simultaneity, it provides a stronger indication of the relevance of each variable.

The impact of *ECI* on the evolution of formal employment was evaluated estimating the following equation, analogous to Equation (13):

$$\log(l)_{i,t} = \alpha + f_i + f_t + \beta_1 ECI_{i,t-1} + \beta_2 ECI_{i,t-1} * log(l)_{i,t-1} + \beta_3 log(l)_{i,t-1} + \beta_i \log(x)_{i,t-1} + \varepsilon \tag{14}$$

where l is the ratio of formal jobs in the economy in relation to population. The interpretation of the variables is the same as in Equation (13).

It is important to note that although the smart diversification score presented in Section 3 is carried out at the industry level, the econometric investigation is carried out at the regional level. Hence, the investigation aims to explore the relationship between economic complexity and regional development. The coefficients of these regressions will then be used to estimate the results of different diversification strategies in Section 5.

Equations (13) and (14) were estimated using region and time fixed effects. Moreover, the system GMM estimator was used to control for the endogeneity of ECI and of the lag of the dependent variable. The method uses lags of the variables in levels and in difference as instruments for the endogenous variables. The identification of the parameters using the system GMM estimator requires both overidentification, tested using Hansen's J test, and no autocorrelation, tested using Arellano and Bond's autoregressive (AR) test.

4.2. Results

The estimation results for Equation (13) are presented in Table 4. Columns (i) to (iv) present the results for municipalities. ECI is always positive and significant. Most importantly, the models corroborate all assumed hypotheses, including very similar parameters for the ECI at the municipality level to those of the seminal study by Hausmann, Hidalgo, et al. (2014) at the country level and using export data. This result is important as it validates the use of employment data in the assessment of economic complexity. The interaction between ECI and GDP per capita is only significant when population is introduced, and its effect, although negative (as expected) is very small. It is interesting to note that population enters with a positive sign, which indicates that at the regional level, increasing the city's population is associated with future increase in its GDP per capita. The lag of GDP per capita is positive, but lower than one, which supports the hypothesis of technological absorption when the log of GDP per capita is used as the dependent variable. The share of employees with tertiary education is not significant. The system GMM regression corroborates the importance of economic complexity as in the fixed effects regressions, but although the Arellano-Bond AR test indicates there is no autocorrelation, Hansen's J test does not suggest overidentification. Hence, these results must be taken with caution. Tertiary education becomes significant, while the coefficients of population and the interaction term change signs.

Columns (v) to (viii) of Table 4 present the results for microregions. The results are similar, but the coefficient of ECI is larger as well as of the interaction with GDP per capita, while population loses its significance except in the system GMM regression. These estimates reinforce the robustness of the results found for municipalities, especialy because now the Arellano-Bond AR test and Hansen's J test suggest the validity of the instruments in the microregion sample.

The findings for Equation (14) are presented in Table 5. Columns (i) to (iv) present the results for municipalities. ECI is not significant in column (i), when introduced along with the interaction of ECI with employment per capita, but when the interaction is removed, it enters with a positive and significant coefficient in columns (ii) and (iii). This is probably due to a high correlation between the two variables. Population enters with a negative sign, which indicates that at the regional level, increasing the city's population is associated with future decrease in employment per capita. The share of employment with tertiary education is positively associated with employment per capita only in the system GMM regression. Yet, the Arellano-Bond AR test and Hansen's J test reject the validity of the instruments.

Columns (v) to (viii) of Table 5 present the results for microregions. The results are once again similar, but the coefficient of ECI is smaller and the interaction with employment per capita enters with a positive coefficient but not significant. Population is still negative and significant, while the share of employment with tertiary education loses its significance. Overall, therefore, it is possible to consider the estimates for microregions more robust. This is not unexpected, since these regions are defined based on the economic interaction within them.[2] The system GMM regression corroborates the importance of economic complexity and once again the associated tests indicate the validity of the instruments. Hence, the more robust system GMM regressions suggest the importance of economic complexity for regional growth.

5. THE CASE OF BELO HORIZONTE

To illustrate how the smart diversification score (*SDS*) can be used to help in devising regional development policies, this section illustrates analysis with the case of the city of Belo Horizonte. The city, located in the state of Minas Gerais, in the centre of Brazil, is among the top 10 cities of the country both in terms of GDP per capita and economic complexity. Moreover, its microregion covers a large industrial area where one of the main automobile factories in Brazil (Fiat autos) is located. Nonetheless, during the period under investigation, it has fallen from the fifth to the ninth position in the ranking of economic complexity of municipalities and microregions in Brazil, which causes concern about the diversification trajectory of the city and stresses the importance of a smart diversification strategy for the region.

Table 4. Economic complexity and GDP per capita.

	Municipalities				Microregions			
Model	**(i) FE**	**(ii) FE**	**(iii) FE**	**(iv) SYS-GMM**	**(v) FE**	**(vi) FE**	**(vii) FE**	**(viii) SYS-GMM**
ECI, lag	0.064**	0.074***	0.073***	0.431***	0.185***	0.177***	0.175***	0.756***
	(0.027)	(0.026)	(0.027)	(0.041)	(0.054)	(0.052)	(0.053)	(0.135)
Log of GDP per capita, lag	0.580***	0.584***	0.582***	0.921***	0.624***	0.635***	0.635***	0.864***
	(0.012)	(0.012)	(0.012)	(0.024)	(0.030)	(0.030)	(0.030)	(0.058)
(ECI * Log of GDP per capita), lag	–0.005	–0.006**	–0.006*	-0.041***	–0.014***	–0.013***	–0.013**	–0.056***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.005)	(0.005)	(0.010)
Log of Population, lag		0.046***	0.041**	–0.020***		0.045	0.045	–0.080***
		(0.017)	(0.017)	(0.007)		(0.047)	(0.047)	(0.034)
Log of Perc. emp. with tert. educ., lag			0.000	0.008***			0.002	0.007
			(0.001)	(0.002)			(0.010)	(0.013)
Constant	3.781***	3.632***	3.661***	1.004***	4.235***	3.972***	3.975***	2.041***
	(0.103)	(0.128)	(0.127)	(0.242)	(0.331)	(0.380)	(0.381)	(0.765)
Observations	61,213	61,213	60,975	27,783	6138	6138	6138	2790
Adjusted R2	0.870	0.870	0.870		0.914	0.914	0.914	
Instruments/lags				19/2-.				19/2-.
Arellano-Bond AR test				0.711				0.941
Hansen J test				0.000				0.170

Note: The dependent variable is the Log of GDP per capita. Robust standard errors in parenthesis. All models are estimated introducing region and year fixed effects. System GMM models regressed using 2-year averages. Significance: * = 10%; ** = 5%; *** = 1%.
Source: Authors' elaboration.

Table 5. Economic complexity and employment per capita.

	Municipalities				Microregions			
Model	**(i) FE**	**(ii) FE**	**(iii) FE**	**(iv) SYS-GMM**	**(v) FE**	**(vi) FE**	**(vii) FE**	**(viii) SYS-GMM**
ECI, lag	−0.086	0.066***	0.058***	0.309**	−0.054	0.039***	0.039***	0.214***
	(0.058)	(0.009)	(0.009)	(0.139)	(0.115)	(0.012)	(0.012)	(0.097)
Log of Employment per capita, lag	0.255***	0.223***	0.230***	1.147***	0.404***	0.348***	0.348***	0.590***
	(0.015)	(0.018)	(0.015)	(0.091)	(0.052)	(0.058)	(0.058)	(0.121)
(ECI * Log of Employment per capita), lag	0.034***				0.014			
	(0.013)				(0.018)			
Log of Population, lag		−0.198***	−0.191***	−0.189**		−0.217***	−0.217***	
		(0.037)	(0.033)	(0.078)		(0.064)	(0.064)	
Log of Perc. emp. with tert. educ., lag			−0.015***	0.040**			0.000	
			(0.005)	(0.016)			(0.011)	
Constant	3.360***	4.011***	3.930***	−0.083	4.056***	5.122***	5.123***	2.875***
	(0.069)	(0.140)	(0.129)	(0.444)	(0.351)	(0.565)	(0.564)	(0.839)
Observations	61,215	61,215	60,977	27,783	6138	6138	6138	2790
Adjusted R2	0.164	0.164	0.166		0.454	0.462	0.462	
Intruments/Lags				11/3				13/2-.
Arellano-Bond AR test				0.055				0.305
Hansen J test				0.008				0.260

Note: The dependent variable is the Log of Employment per capita. Robust standard errors in parenthesis. All models are estimated introducing region and year fixed effects. System GMM models regressed using 2-year averages. Significance: * = 10%; ** = 5%; *** = 1%.
Source: Authors' elaboration.

5.1. Identifying promising activities for Belo Horizonte

Figure 2 shows the *Activity Space* of the municipality and of the microregion of Belo Horizonte in 2018. The coloured dots mark the activities in which the region is competitive, i.e., with $RCA > 1$. In this figure it is possible to identify four clusters: (1) public services; (2) modern services; (3) trade; (4) manufacturing; (5) construction.

Using the *SDS* presented in Section 3.3 it is possible to identify the top 10 most promising activities within each the diversification strategy presented in Section 3.4. Figures 2C and D show the position of these 40 promising activities in the *Activity Spaces*. Tables A1 and A2, in the Appendix in the online supplemental data, present the list of activities in each of the four subgroups of Table 3. The figure shows that activities in the *Strategic Bets* (in yellow) strategy are in general associated with industrial activities, which tend to present higher complexity, while the remaining activities identified are spread around the network.

Figure 2 also illustrates that some of the activities identified as promising for the municipality of Belo Horizonte are already competitive in its microregion, such as *Manufacture of measurement, test and control equipment* (CNAE 26515). This highlights the importance of further investigation into what is the adequate regional level of analysis, since using different levels will lead to different prescriptions. Moreover, this finding stresses the importance of considering regional interactions, since the decision to foster one activity might be hampered by the existence of a strong competitor nearby. Furthermore, it might be counterproductive for nearby cities to incentivise the same activity simultaneously.

To identify promising macro-sectors, it is also possible to group promising activities into more aggregated CNAE sectors, at two digits, as shown in Figure 3. Among the promising activities, the macro-sectors *Machines*, *Electronics* and *Vehicles* encompass eight activities. Among these activities, six are from the *Bets* strategy. Next comes the *Chemical-Pharmaceutical* (three activities) sector, and *Metallurgy* (two activities). Moreover, several specialised services

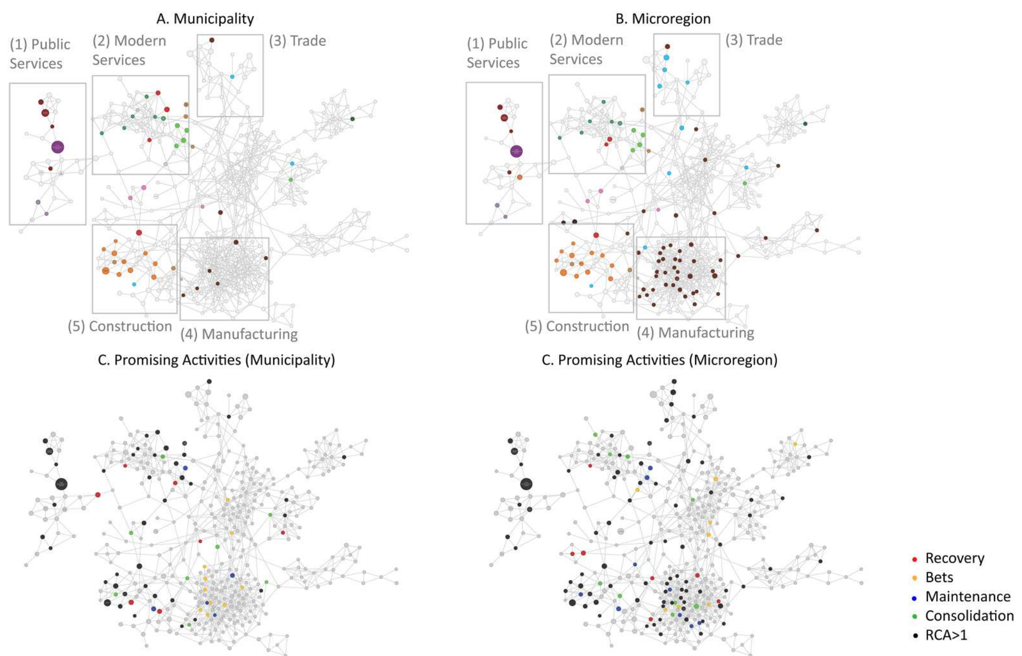

Figure 2. Activity space of Belo Horizonte in 2018
Source: Authors' elaboration.

associated with production are identified as promising, such as *Infrastructure*, *IT* and *Machine Maintenance* (six activities). The results are very similar both for the municipality and the microregion. The message that emerges from this analysis of promising sectors classified into macro-sectors, therefore, is that the *Machinery*, *Chemicals* and *Metal-Mechanical* sectors are the ones that concentrate the greatest number of promising activities. More generally, the importance of developing *High-Tech Manufactures* and *Modern Services* linked to production is also highlighted.

5.2. Projecting the diversification gains of Belo Horizonte

Using the estimates of the impact of *ECI* on the growth rates of GDP per capita and employment presented in Section 4 it is possible to estimate by how much would these variables increase in response to the increase in the city's *ECI* generated by becoming competitive in the industries identified as promising.

Table 6 presents the average marginal effects of different diversification strategies. Notice that the number of activities is not equal to 10 in all subgroups. In the case of *Consolidation*, these are the activities in this subgroup that do not have $RCA > 1$, while in the case of *Recovery*, these are the activities in this subgroup that have $RCA > 1$ (see Appendixes in the online supplemental data). We devote special attention to the *Consolidation* and *Bets* strategies, and their combination. The *Recovery* strategy is presented to show the inverse relationship, measuring the negative effects of losing *RCA* in these activities.[3] The estimated effects of *RCA* loss in *Recovery* activities indicate that losing competitiveness in these seven sectors could lead to a 3.41% decrease in the municipal *ECI*, which would result in reductions of 0.57 percentage points in the municipality's growth rate and of 0.22 percentage points in the growth rate of the formal employment ratio. At the other end of the analysis, the acquisition of *RCA* in the 10 products of the *Bets* strategy would bring a 7.9% increase in *ECI* and an increase in the municipal GDP per capita growth rate of 1.3 percentage points. The effect in the formal labour market is also considerable, increasing the formal employment growth rate by almost 0.5 percentage points.

The results presented in this section demonstrate the importance of devising short-, medium- and long-term development strategies. The results show that the opportunity cost for the municipality of Belo Horizonte is considerably higher than that of the microregion in all the strategies presented, something to be carefully looked at by local public managers.

Finally, it is important to note that the increase in complexity has effects not only on growth, but also on exports (Romero & Britto, 2017), on inequality (Hartmann et al., 2017) and on greenhouse gas emission (Romero & Gramkow, 2021). The combination of this evidence, therefore, supports the need for smart diversification strategies to overcome the bottlenecks faced by peripheral regions, reinforcing the importance of structural change for economic development.

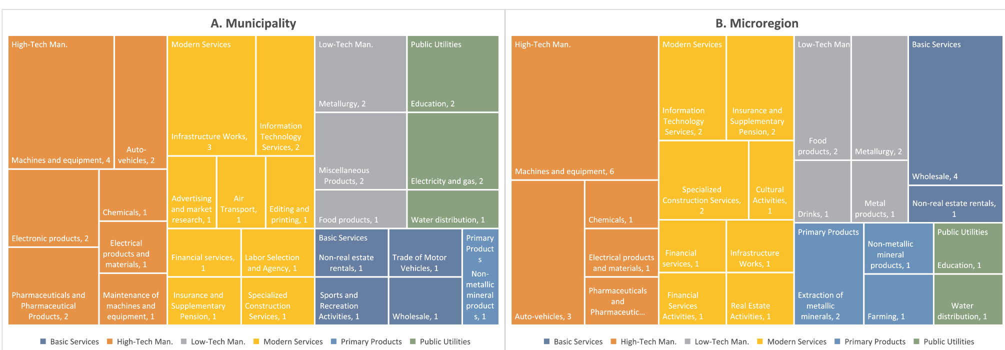

Figure 3. Promising macro-sectors for Belo Horizonte in 2018
Source: Authors' elaboration.

Table 6. Projected gains in GDP per capita and employment by acquiring RCA in the indicated promising activities: municipality and microregion of Belo Horizonte.

	Municipality				Microregion			
	Recovery*	Consolidation	Strategic Bets	Consolidation + Bets	Recovery*	Consolidation	Strategic Bets	Consolidation + Bets
Total products in the category	7	5	10	15	9	4	10	14
Projected ECI	2.741	2.901	3.062	3.113	2.836	2.990	3.078	3.176
ECI gain (%)	−3.41%	2.23%	7.90%	9.70%	−1.59%	3.75%	6.81%	10.22%
Rate of change of GDP per capita (%)	−0.70%	0.46%	1.63%	2.00%	−0.21%	0.50%	0.90%	1.35%
Rate of change of employment (%)	−15.03%	9.84%	34.78%	42.71%	−2.55%	6.03%	10.94%	16.41%
Counterfactual level of employment (gain)**	−0.22%	0.14%	0.50%	0.61%	−0.17%	0.40%	0.73%	1.09%
Counterfactual GDP per capita (gain)***	−0.57%	0.37%	1.32%	1.63%	−0.31%	0.73%	1.32%	1.98%

Notes: The ECI for Belo Horizonte in 2019 was 2.84 and for the microregion 2.88. The average marginal effect of ECI on GDP growth was estimated at 0.073 and on employment growth at 0.557. * The recovery strategy shows the loss if the municipality loses RCA in the sectors. ** Counterfactual formal employment level was calculated from the average marginal effect of ECI on GDP per capita estimated at 0.02 for the municipality and 0.04 for the region. *** Counterfactual GDP per capita was calculated from the average marginal effect of ECI on GDP per capita estimated at 0.06 for the municipality and 0.07 for the region.
Source: Authors elaboration based on data from RAIS and IBGE.

The challenge is to create the right incentives for economic diversification to take place in the most promising direction.

6. CONCLUDING REMARKS

This paper sought to contribute to the literature on economic complexity and regional development in three ways: by (i) conceiving of a new method – smart diversification score (SDS) – to be used by policymakers to rank promising activities for short-, medium- and long-term diversification, (ii) testing its validity in the municipal and microregion levels and (iii) exemplifying its use for policy purposes. The SDS method combines a series of indicators, as proposed by Hausmann et al. (2017), but uses weights estimated using principal component analysis.

After reporting the positive impact of regional complexity, calculated using employment data, on employment and GDP per capita growth for Brazilian municipalities and microregions, the paper has assessed the potential of SDS to increase the ECI at the municipal and microregion levels. Looking backwards, i.e., using data from 2007 and 2018, the paper found that SDS can predict up to 39.4% of the diversification activities that happened in 1033 Brazilian municipalities that experienced increases in ECI during the period. Looking forward, the paper calculated the SDS for the city of Belo Horizonte, suggesting a balanced portfolio of related and unrelated activities for diversification and estimated the potential economic gains to be obtained following different development paths. As expected, the results suggest substantial gains to be obtained from the proposed diversification strategies.

The evidence presented in this paper, therefore, provides important contributions for the formulation of regional development policies. More specifically, it provides an interesting framework for identifying promising activities to be candidates to be targeted by policymakers to promote regional economic development. Moreover, the investigation presented in the paper highlights the importance of defining the appropriate regional level for the policies in order to avoid unproductive competition between neighbouring regions.

Nonetheless, research is still required to improve even further the methods used to identify promising activities for regional diversification. Although the framework outlined in this paper contributes to the definition of short-, medium- and long-term diversification strategies, associated with related and unrelated activities, it is still necessary to investigate the right proportion between each of these types of activities, taking into account the complexity level of each region (Hidalgo, 2023). Moreover, further research is still required to determine more precisely the key variables that should be taken into account to identify promising activities. And finally, there is still considerable room to explore the specific policy measures that should be taken to foster the development of the activities identified as promising in each territory.

ACKNOWLEDGEMENTS

The authors would like to thank the two anonymous referees for their helpful comments.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

FUNDING

Funding from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [grant numbers 313755/2021-3, 311052/2022-3 and 309982/2015-4], from Fundação de Amparo à

Pesquisa do Estado de Minas Gerais (FAPEMIG) [grant numbers APQ-02518-17 and 04746-22], and from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [grant number 88887.837602/2023-00], is gratefully acknowledged.

NOTES

1 This network can also be found and explored in the DataViva platform, at www.dataviva.info/en/.
2 The regressions presented in Tables 4 and 5 were also performed using clustered standard errors, but the results did not change significantly. The results are available from the authors upon request.
3 *Maintenance* is not considered an actual strategy, since these sectors are well developed in either the municipality/microregion of Belo Horizonte. That is, the model assumes that these sectors are not under risk, they were shown to illustrate the most competitive products and activities of Belo Horizonte.

ORCID

João P. Romero http://orcid.org/0000-0003-1254-0074
Elton Freitas http://orcid.org/0000-0001-8887-4471
Fabrício Silveira http://orcid.org/0000-0003-1240-3942
Gustavo Britto http://orcid.org/0000-0002-5285-3684
Fernanda Cimini http://orcid.org/0000-0003-2170-6443
Frederico G. Jayme Jr. http://orcid.org/0000-0002-4617-0107

REFERENCES

Almeida, P., & Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, *45*(7), 905–917. https://doi.org/10.1287/mnsc.45.7.905

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), 1252–1268. https://doi.org/10.1080/00343404.2018.1437900

Balland, P. A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O'Clery, N., & Rigby, D. (2022). The new paradigm of economic complexity. *Research Policy*, *51*(3), 104450. https://doi.org/10.1016/j.respol.2021.104450

Bishop, P., & Gripaios, P. (2009). Spatial externalities, relatedness and sector employment growth in Great Britain. *Regional Studies*, *44*(4), 443–454. https://doi.org/10.1080/00343400802508810

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

Boschma, R., & Frenken, K. (2011). The emerging empirics of evolutionary economic geography. *Journal of Economic Geography*, *11*, 295–3017.

Boschma, R., & Iammarino, S. (2009). Related variety, trade linkages and regional growth in Italy. *Economic Geography*, *85*(3), 289–311. https://doi.org/10.1111/j.1944-8287.2009.01034.x

Britto, G., Romero, J. P., Freitas, E., & Coelho, C. (2019). The great divide: Economic complexity and development paths in Brazil and the Republic of Korea. *Cepal Review*, *2019*(127), 191–213. https://doi.org/10.18356/d06fff2f-en

Buyukyazici, D., Mazzoni, L., Riccaboni, M., & Serti, F. (2023). Workplace skills as regional capabilities: Relatedness, complexity and industrial diversification of regions. *Regional Studies*, *58*(3), 469–489. https://doi.org/10.1080/00343404.2023.2206868

Chávez, J. C., Mosqueda, M. T., & Gómes-Zaldívar, M. (2017). Economic complexity and regional growth performance: Evidence from the Mexican economy. *Review of Regional Studies*, *47*(2), 201–219. https://doi.org/10.52324/001c.8023

Davies, B., & Maré, D. (2021). Relatedness, complexity and local growth. *Regional Studies*, *55*(3), 479–494. https://doi.org/10.1080/00343404.2020.1802418

De Groot, H. L. F., Poot, J., & Smit, M. J. (2009). Agglomeration externalities, innovation and regional growth: theoretical perspectives and meta-analysis. In R. Capello, & J. Nijkamp (Eds.), *Handbook of Regional Growth and Development Theories* (pp. 256–281). Edward Elgar Pub.

Essletzbichler, J. (2005). Diversity, stability and regional growth in the United States, 1975-2002. *Papers in Evolutionary Economic Geography*, n. 0513.

Farjoun, M. (1994). Beyond industry boundaries: human expertise, diversification and resource-related industry groups. *Organization Science*, *5*(2), 185–199. https://doi.org/10.1287/orsc.5.2.185

Freitas, E. (2019) Indústrias relacionadas, complexidade econômica e diversificação regional: uma aplicação para microrregiões brasileiras, PhD Thesis, Center for Regional Development and Planning (Cedeplar), Universidade Federal de Minas Gerais (UFMG).

Frenken, K., Van Oort, F. G., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, *41*(5), 685–697. https://doi.org/10.1080/00343400601120296

Furtado, C. (1964). *Development and Underdevelopment*. University of California Press.

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*, (6), 1355–1377.

Gao, J., Jun, B., Pentland, A. S., Zhou, T., & Hidalgo, C. A. (2021). Spillovers across industries and regions in China's regional economic diversification. *Regional Studies*, *55*(7), 1311–1326. https://doi.org/10.1080/00343404.2021.1883191

Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G., & Van Den Oord, A. (2007). Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research Policy*, *37*(10), 1717–1731. https://doi.org/10.1016/j.respol.2008.08.010

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Schleifer, A. (1992). Growth in cities. *Journal of Political Economy*, *100*(6), 1126–1152. https://doi.org/10.1086/261856

Hartmann, D., Bezerra, M., & Pinheiro, F. L. (2019). Identifying smart strategies for economic diversification and inclusive growth in developing economies. The Case of Paraguay. *SSRN Electronic Journal*, *8*, 1–41. https://doi.org/10.2139/ssrn.3346790

Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristaran, M., & Hidalgo, C. A. (2017). Linking economic complexity, institutions and income inequality. *World Development*, *93*, 75–93. https://doi.org/10.1016/j.worlddev.2016.12.020

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*, 1–29. https://doi.org/10.1016/j.strueco.2021.04.007

Hausmann, R., Barrios, D., Muhaj, D., Noor, S., Pan, C., Santos, M. A., Tapia, J., & Zuccolo, B. (2020b). Emerging cities as independent engines of growth: the case of Buenos Aires. CID Working Paper, No. 385.

Hausmann, R., & Chauvin, J. (2015). Moving to the adjacent possible: Discovering paths of export diversification in Rwanda. In *Center for International Development (CID) Faculty Working Paper, n. 294*. Harvard University.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Chung, S., Jimenez, J., Simões, A., & Yildirim, M. A. (2014). *The atlas of economics complexity – mapping paths to prosperity*. Puritan Press. p. 364.

Hausmann, R., Pietrobelli, C., & Santos, M. A. (2020a). Place-specific determinants of income gaps: new subnational evidence from Mexico. *CID Working Paper*, No. 343.

Hausmann, R., Santos, M. A., & Obach, J. (2017). Appraising the economic potential of panama: policy recommendations for sustainable and inclusive growth. In *Center for international development (CID) faculty working paper*, n. 334. Harvard University.

Henderson, J. V., Kuncoro, A., & Turner, M. (1995). Industrial development in cities. *Journal of Political Economy*, *103*(5), 1067–1085. https://doi.org/10.1086/262013

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, *3*(2), 92–113. https://doi.org/10.1038/s42254-020-00275-1

Hidalgo, C. A. (2023). The policy implications of economic complexity. *Research Policy*, *52*(9), 1–17. https://doi.org/10.1016/j.respol.2023.104863

Hidalgo, C., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, *106*(26), 10570–10575. https://doi.org/10.1073/pnas.0900943106

Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, *317*(5837), 482–487. https://doi.org/10.1126/science.1144581

Hirschman, A. O. (1958). *The strategy of economic development*. Yale University Press.

Kaldor, N. (1966). *Causes of the slow rate of economic growth of the United Kingdom*. Cambridge University Press.

Li, Y., & Rigby, D. (2023). Relatedness, complexity and economic growth in Chinese cities. *International Regional Science Review*, *46*(1), 3–37. https://doi.org/10.1177/01600176221082308

Lo Turco, A., & Maggioni, D. (2022). The knowledge and skill content of production complexity. *Research Policy*, *51*(8), 104059. https://doi.org/10.1016/j.respol.2020.104059

Lundvall, B. A.1992). *National systems of innovation: Towards a theory of innovation and interactive learning*. Pinter.

Martin, R., & Sunley, P. (2006). Path dependence and regional economic evolution. *Journal of Economic Geography*, *6*(4), 395–437. https://doi.org/10.1093/jeg/lbl012

Mealy, P., & Teytelboym, A. (2021). Economic complexity and the green economy. *Research Policy*, *51*(8), 1–24. https://doi.org/10.1016/j.respol.2020.103948

Nelson, R. R.1993). *National innovation systems: A comparative analysis*. Oxford University Press.

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

Pinheiro, F. L., Balland, P. A., Boschma, R., & Hartmann, D. (2022a). The dark side of the geography of innovation: relatedness, complexity and regional inequality in Europe. *Regional Studies*, 1–16. https://doi.org/10.1080/00343404.2022.2106362

Pinheiro, F. L., Hartmann, D., Boschma, R., & Hidalgo, C. A. (2022b). The time and frequency of unrelated diversification. *Research Policy*, *51*(8), 104323. https://doi.org/10.1016/j.respol.2021.104323

Prebisch, R. (1962). The economic development of Latin America and its principal problems. *Economic Bulletin for Latin America*, *7*(1), 1–19.

Queiroz, A., Romero, J. P., & Freitas, E. (2023). Economic complexity and employment in Brazilian states. *Cepal Review*, *2023*(139), 177–196. https://doi.org/10.18356/16840348-2023-139-9

Romero, J. P., & Britto, G. (2017). Increasing returns to scale, technological catch-up and research intensity: Endogenising the Verdoorn coefficient. *Cambridge Journal of Economics*, *41*(2), 391–412. https://doi.org/10.1093/cje/bew030

Romero, J. P., & Freitas, E. (2018). Setores promissores para o desenvolvimento do Brasil: complexidade e espaço do produto como instrumentos de política. In M. Viegas, & E. Albuquerque (Eds.), *Alternativas para uma crise de múltiplas dimensões* (1 ed, pp. 358–374). Cedeplar-UFMG.

Romero, J. P., & Gramkow, C. (2021). Economic complexity and greenhouse gas emissions. *World Development*, *139*, 1–18. https://doi.org/10.1016/j.worlddev.2020.105317

Romero, J. P., & Silveira, F. (2019) Mudança estrutural e complexidade econômica: identificando setores promissores para o desenvolvimento dos estados brasileiros, p. 137-160, in: Chialliato-Leite, M. V. (org.), Alternativas para o desenvolvimento brasileiro: novos horizontes para a mudança estrutural com igualdade, Santiago: Comisión Económica para América Latina y el Caribe (CEPAL).

Schumpeter, J. (1934). *The theory of economic development*. Harvard University Press.

Scott, A. J. (1988). *New industrial spaces: Flexible production organization and regional development in North America and Western Europe*. Pion.

Simões, R. F., & Freitas, E. F. (2014). Urban attributes and regional differences in productivity: Evidence from the external economics of Brazilian micro-regions from 2000–2010. *Journal of Economic & Financial Studies*, *2*(1), 30–44. https://doi.org/10.18533/jefs.v2i01.51

Storper, M., & Walker, R. (1989). *The capitalist imperative: Territory, technology and industrial growth*. Basil Blackwell.

Teixeira, F. O., Missio, F. J., & Dathein, R. (2022). Economic complexity, structural transformation and economic growth in a regional context: Evidence for Brazil. *PSL Quarterly Review*, *75*(300), 47–79. https://doi.org/10.13133/2037-3643/17505