Universidade Federal de Minas Gerais
Faculdade de Ciências Econômicas
Centro de Desenvolvimento e Planejamento Regional

Arthur Ribeiro Queiroz

Economic Complexity and Regional Inequalities: good news for some, bad news for others

Complexidade Econômica e Desigualdades Regionais: uma boa notícia para alguns, uma má notícia para os demais

Belo Horizonte
2023

Arthur Ribeiro Queiroz

Economic Complexity and Regional Inequalities: good news for some, bad news for others

Dissertação apresentada ao curso de Mestrado do Centro de Desenvolvimento e Planejamento Regional da Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais como requisito parcial à obtenção do Título de Mestre em Economia.

Universidade Federal de Minas Gerais

Orientador: Prof. Dr. João Prates Romero
Co-orientador: Prof. Dr. Elton Eduardo Freitas

Belo Horizonte
2023

Resumo

Esta dissertação investiga o fenômeno da diversificação relacionada nas microrregiões brasileiras, com foco específico em sua natureza desigual. A pesquisa destaca que a relatedness não apenas impulsiona o processo de diversificação, mas também desempenha um papel significativo na intensificação da desigualdade regional. O Capítulo 1 estabelece as bases teóricas que definem a Geografia Econômica como uma Ciência Evolucionária (BOSCHMA; FRENKEN, 2006). Essa perspectiva enfatiza que o desenvolvimento econômico regional resulta de mudanças estruturais que ocorrem ao longo do tempo em vários níveis, como empresas, setores e instituições. Com base nesse quadro, a dissertação argumenta que a diversificação relacionada pode intensificar as desigualdades entre as regiões devido à dependência de trajetória, que restringe a produção de setores complexos apenas às regiões que já possuem complexidade. O Capítulo 2 investiga empiricamente a influência da relatedness e da complexidade na entrada e saída de empresas nas regiões, considerando diferentes níveis de complexidade. A pesquisa constata que as regiões têm maior probabilidade de iniciar a produção em setores relacionados ao seu portfólio existente e maior probabilidade de interromper a produção em setores menos relacionados. Além disso, o nível de complexidade regional tem um impacto notável no processo de diversificação. Em regiões altamente complexas, maior complexidade setorial aumenta a probabilidade de entrada, mas tem impacto mínimo sobre a probabilidade de saída. Por outro lado, em regiões menos complexas, a maior complexidade do setor diminui a sua probabilidade de entrada e contribui significativamente para a sua saída. O Capítulo 3 examina as implicações desse processo na criação de empregos regionais. A pesquisa adapta a metodologia de Moretti and Thulin (2013) para mensurar multiplicadores de emprego locais para setores complexos e não complexos. Os resultados indicam que o setor complexo tem multiplicadores de emprego mais altos. No entanto, os multiplicadores variam de acordo com a complexidade das regiões, com os multiplicadores do setor complexo sendo mais expressivos em regiões já complexas. A dissertação conclui destacando as implicações dos resultados da pesquisa. A natureza desigual da diversificação relacionada exige políticas direcionadas para abordar as disparidades de desenvolvimento entre as regiões. Os formuladores de políticas devem criar oportunidades para que as regiões menos complexas desenvolvam as capacidades necessárias para a diversificação em setores complexos, rompendo assim esse ciclo de desenvolvimento retrógrado.

Palavras-chave: Relatedness, Complexidade, Diversificação, Multiplicadores de Emprego.

Abstract

This master's thesis investigates the phenomenon of related diversification in Brazilian micro-regions, with a specific focus on its unequal nature. The research highlights that relatedness not only drives the diversification process but also plays a significant role in exacerbating regional inequality. Chapter 1 establishes the theoretical foundations that define Economic Geography as an Evolutionary Science (BOSCHMA; FRENKEN, 2006). This perspective emphasizes that regional economic development results from structural changes occurring over time at various levels, such as firms, sectors, and institutions. Based on this framework, the thesis argues that related diversification can intensify inequalities between regions due to path dependence, which restricts the production of complex sectors to regions already possessing complexity. Chapter 2 empirically investigates the influence of relatedness and complexity on the entry and exit of firms in regions, considering different levels of complexity. The research finds that regions are more likely to start producing in sectors related to their existing portfolio and more likely to stop producing in less related sectors. Furthermore, the level of regional complexity has a notable impact on the diversification process. In highly complex regions, greater sectoral complexity increases the probability of entry but has minimal impact on the probability of exit. On the other hand, in less complex regions, the greater complexity of the sector decreases the probability of entry and contributes significantly to its exit. Chapter 3 examines the implications of this process on regional job creation. The research adapts the methodology of Moretti and Thulin (2013) to measure local employment multipliers for complex and non-complex sectors. The results indicate that the complex sector has higher employment multipliers. However, the multipliers vary according to the complexity of the regions, with complex sector multipliers being more expressive in already complex regions. The thesis concludes by emphasizing the political implications of the research results. The uneven nature of related diversification calls for targeted policies to address development disparities across regions. Policymakers must create opportunities for less complex regions to develop the capabilities needed to diversify into complex sectors, thus breaking the retrograde development cycle.

Keywords: Relatedness, Complexity, Diversification, Employment Multipliers.

List of Figures

List of Tables

Contents

Introduction

Regional growth can be explained by the diversity of sectors within the local economy (GLAESER et al., 1992; JACOBS, 1969). The presence of diverse skills, knowledge, and production techniques for existing goods shortens the path to producing new goods and stimulates innovation opportunities. However, the process of regional economic diversification is complex and influenced by various dynamics. For instance, economies may diversify into related or unrelated sectors (FRENKEN; OORT; VERBURG, 2007), each with distinct implications: the former supports employment growth, while the latter prevents unemployment growth. Nevertheless, diversification towards related sectors is more common, with empirical studies providing evidence for this (HIDALGO et al., 2018). Hence, it can be asserted that relatedness plays a crucial role as a driver of regional diversification (BOSCHMA, 2017).

The economic complexity approach has bolstered this interpretation by providing measurements of crucial concepts for understanding regional diversification (HIDALGO et al., 2007). In essence, the complexity of regions gauges the diversity of their capabilities. In other words, more complex regions possess a broader array of components (skills, knowledge, techniques) to assemble a greater variety of products, whereas this variety is limited in less complex regions. As paraphrased from Hidalgo and Hausmann (2009), the disparity in productivity between regions is determined by this range of internal capabilities. Moreover, through measuring the interconnections among products and between products and the local productive structure, this approach has provided empirical evidence that supports the thesis that diversification predominantly occurs towards related sectors.

Nonetheless, economic geography should be understood as an evolutionary science (BOSCHMA; FRENKEN, 2006). This implies that current economic conditions are fundamentally shaped by pre-existing factors, leading to local development characterized by path dependence. Consequently, relatedness, besides being a driver of diversification, also plays a significant role in regional inequalities. The trend of economies to diversify in a related manner widens the gap between more and less complex regions, as diversification into complex sectors is limited to already complex regions. As a result, complex regions diversify into more inclusive and knowledge-intensive sectors, while less complex regions diversify into resource-based sectors with a less skilled workforce (HARTMANN; PINHEIRO, 2022). This process elucidates why relatedness is *good news* for some and *bad news* for others.

The impossibility of achieving structural changes towards a more complex economy places regions in a challenging situation, given that complexity is a positive predictor of future economic performance (HAUSMANN et al., 2014) and employment (ROMERO et al., 2022; QUEIROZ; ROMERO; FREITAS, 2023). Consequently, incorporating complexity to assess regional economic diversification reestablishes the concept of circular cumulative causation proposed by Myrdal (1957). This means that the forces of related diversification perpetuate a feedback loop that further hinders less complex regions from becoming more complex (PINHEIRO et al., 2022).

Although this process is fundamental for comprehending regional economic development, there are few contributions in the literature that have focused on it. Moreover, understanding how this phenomenon manifests itself in Brazil is essential, given its highly unequal nature and widely diverse regions. Therefore, the primary objective of this master's thesis is to assess the uneven character of related diversification in Brazilian regions. To achieve this, in addition to the Introduction and Conclusions, the thesis is structured into three chapters. The first chapter presents a comprehensive discussion and literature review, while the subsequent chapters offer original empirical contributions.

Chapter 1 has the primary objective of reviewing the literature that examines the relationship among diversification, complexity, and inequality from a regional perspective. To achieve this, three specific goals guide the discussion. Firstly, the chapter aims to establish the rationale behind viewing Economic Geography as an Evolutionary Science. This exploration serves as a foundation for the second goal: comprehending how the literature explains the process of regional economic diversification. Lastly, the chapter seeks to theoretically demonstrate that related diversification also acts as a driver of regional inequalities.

The discussion organized in this manner facilitates the identification of issues not yet addressed by the literature, making it a significant contribution of Chapter 1. Among the highlighted gaps, the absence of answers to specific questions stands out. For instance, what is the impact of sectoral complexity on the probability of a firm entering or exiting the local portfolio? How does this effect vary with different levels of regional complexity? What are the implications of diversifying into more complex sectors? Additionally, how does increased complexity influence key factors in local economies, such as job creation? These questions served as guiding principles for the empirical research conducted in the subsequent chapters.

Chapter 2 aims to empirically investigate why relatedness is good news for some regions and bad news for others. To achieve this, the chapter prioritizes examining the impact of economic complexity on the diversification process. In practical terms, the objective is to assess the entry and exit of firms from local portfolios, with a focus on sectoral and regional complexity. To structure our model, we build upon the work of Neffke, Henning and Boschma (2011). Additionally, we classify regions into four groups based on their complexity levels: Low, Medium-Low, Medium-High, and High. This segmentation enables the analysis of hypotheses that have not been fully explored in prior studies.

Using data from the Brazilian formal labor market and employing three different estimation strategies (OLS, Logit, and Probit), Chapter 2 makes significant contributions to the literature. Beyond confirming the common hypotheses that relatedness increases the likelihood of firm entry and reduces the likelihood of firm exit, the chapter draws two main conclusions:

i) In highly complex regions, an increase in sectoral complexity increases the probability of entry of new sectors, while having a minimal impact on the probability of exit of sectors.

ii) In less complex regions, greater sectoral complexity decreases the probability of entry and significantly contributes to sector exit.

Chapter 3 aims to answer the question of how good or how bad the news is in terms of one variable: employment. The practical objective is to understand the heterogeneity of local employment multipliers based on variations in regional complexity. To achieve this, we adapted Moretti and Thulin (2013)'s conceptual framework, distinguishing the economy into a complex sector and a non-complex sector, and then analyzed the multiplier effect of both sectors. For example, we examined how many jobs are generated in the non-complex sector for each additional job created in the complex sector. All potential relationships between them were thoroughly tested.

Utilizing the same data as in Chapter 2 and maintaining the segmentation of regions by complexity, this novel contribution to the literature yielded the following evidence:

i) The complex sector is more effective in generating jobs.

ii) The bad news for less complex regions is a complex sector unable to generate jobs.

iii) The good news for complex regions is a complex sector capable of generating between 1.06 and 1.46 jobs in the same sector and between 1.71 and 3.25 in the non-complex sector for each additional job.

The final section of this master's thesis presents the primary research findings, discusses their contributions to the literature on regional economic diversification and its implications, particularly in terms of policies. Additionally, the section addresses the main limitations of the study and highlights new research avenues based on the thesis findings.

1 Diversification, Complexity, and Inequalities: a Regional Review

1.1 Introduction

The process of economic development has the productive structure as one of its most important determinants. For this reason, classical theorists of economic development argue that structural change is a key factor for achieving better future economic performance (SCHUMPETER, 1934; LEWIS, 1955; HIRSCHMAN, 1958; MYRDAL, 1957; PREBISCH, 1950; FURTADO, 1964). Therefore, differences in the composition of the local productive structure are sources of differentiation between regions.

Internal structural conditions separate underdeveloped from developed countries. Furtado (1964), one of the main formulators of ECLAC's[1] economic theory, characterizes underdevelopment as the penetration of modern capitalist companies into the midst of archaic structures. This process reveals the inability of the productive structure of underdeveloped countries to incorporate technical progress, creating limiting bottlenecks for economic development. Such conditions support the division of countries between the center and periphery. While the composition of the productive structure of central countries is dominated by the manufacturing sector, that of peripheral countries is marked by resource-based sectors (FURTADO, 1964; PREBISCH, 1950). This structural difference between countries gained new empirical substance with the approaches inaugurated by Hidalgo et al. (2007) and Hidalgo and Hausmann (2009).

ECLAC's center-periphery logic was resumed by the economic complexity approach. Hidalgo et al. (2007) used international trade data to construct a Product Space, a network of relationships between products exported worldwide, aiming to explain structural differences between countries. The Product Space connects products based on the possibility of being competitively co-produced, resulting in clusters of products that require similar production capabilities. At the center of the network are the most sophisticated products, while less sophisticated ones are found at the peripheral areas. Hidalgo et al. (2007) demonstrated that industrialized countries have a productive structure characterized by sophisticated products, whereas countries in Latin America, for instance, are specialized in products located at the periphery of the network.

Furthermore, Hidalgo and Hausmann (2009) propose that a country's level of productivity is determined by the diversity of its internal capabilities. They argue that income differences between countries are primarily explained by variations in economic complexity, which essentially measures the diversity of capabilities a country possesses. To represent this concept, they combine two fundamental factors: the diversification of coun-

[1] Economic Comission for Latin America and the Caribbean (ECLAC).

tries, measured by the number of products they export competitively, and the ubiquity of products, measured by the number of countries that export them competitively. The interaction of these characteristics allows for an assessment of the variety of capabilities that exist in a country and are required for production in a specific sector (HAUSMANN et al., 2014). Based on this interaction, Hidalgo and Hausmann (2009) develop two crucial measures: the Economic Complexity Index (ECI) and the Product Complexity Index (PCI). These indicators play a significant role in supporting the discussions and analyses undertaken during this period.

In parallel with the evolution of the economic complexity literature, there has been a resurgence in the discussion about the significance of the productive structure in regional economic development, fueled by new assessments of the effects of Jacobs (1969) and Marshall (1920) externalities. The central question that emerged from this debate was whether firms benefit more from proximity to other firms in the same industry or from being close to firms in different industries. In essence, this literature aims to comprehend the primary source of knowledge spillovers – either through diversification (Jacobs) or specialization (Marshall). The seminal work of Glaeser et al. (1992) provided critical evidence for this debate, revealing that employment growth in the industries studied was more closely associated with urban diversification and local competition than with regional specialization. Subsequently, numerous other studies have followed suit, designating the diversification process as the central element of investigation.

Frenken, Oort and Verburg (2007) and Boschma and Iammarino (2009) argue that, in addition to the dichotomy presented above, it is necessary to delve deeper into Jacobs (1969)' concept of diversification. According to them, the environment conducive to knowledge spillovers is one where diversification occurs in sectors that share related skills and capabilities. Building on this idea, a significant body of literature has identified that regions tend to diversify into activities that are related to the existing activities in the region (NEFFKE; HENNING; BOSCHMA, 2011; BOSCHMA; MINONDO; NAVARRO, 2013; BOSCHMA; BALLAND; KOGLER, 2015; RIGBY, 2015; HIDALGO et al., 2018). The observation that regions exhibit a propensity to diversify into related sectors has been synthesized as the Principle of Relatedness (HIDALGO et al., 2018). This path dependence in regional development holds significant implications.

Pinheiro et al. (2022) examined the consequences of this diversification dynamic and labeled it as "the dark side of the geography of innovation". They posit that wealthy and complex regions tend to diversify into complex and related activities, while economically disadvantaged regions are limited to low complexity activities. As a result, related diversification increases the existing disparity between regions, contributing to regional inequality. This is why relatedness is *good news for some* and *bad news for others*. Furthermore, this process demonstrates how the production of complex goods is concentrated in space.

Based on these findings, it is evident that the most complex activities are concentrated in areas with a higher division of productive knowledge. As stressed by Balland and Rigby (2017), the distribution of knowledge complexity is uneven across space. Consequently, the evidence presented by Balland et al. (2020) reveals that the level of complexity of each sector influences its spatial distribution. This supports the central conclusion of the authors: complex activities tend to be concentrated in large urban centers, as they require a deeper division of knowledge that is distributed among many actors. As a result, scholars highlight that an increase in economic complexity may be associated with a growth in spatial inequality.

This chapter aims to provide a comprehensive review of the literature that connects regional complexity with inequalities. By examining the works that characterize the regional economic diversification process and underscore the significance of related diversification, the chapter advances the argument that analyzing related diversification through the lens of complexity reveals that relatedness is not merely a driver of the regional diversification process, but also a driver of regional inequalities. This objective lays the foundation for the discussions in Chapters 2 and 3 and, as its primary contribution, it highlights issues that still require attention within the literature.

To this end, the chapter is structured as follows. Section 2 describes the theoretical framework for defining economic geography as an evolutionary science, presenting key concepts for understanding economic diversification. Section 3 focuses on the literature that discusses regional economic diversification. First, the role of agglomeration economies is discussed, contrasting specialization and diversification. Second, relatedness is emphasized as central to diversification. Third, it examines the idea that relatedness, in addition to promoting diversification, also exacerbates regional inequalities. Fourthly, works that are based on the discussed concepts to propose policies guided by economic diversification are presented. Finally, section 4 provides the concluding remarks of the chapter.

1.2 Economic Geography as an Evolutionary Science

The existence and persistence of agglomeration of productive activities are explained in different ways in the economic literature. Boschma and Frenken (2006) assess similarities and dissimilarities between New Economic Geography (NEG), Institutional Economic Geography (IEG), and Evolutionary Economic Geography (EEG). They adapt Veblen (1898)'s[2] central question to highlight the evolutionary character of economic geography.

[2] Veblen (1898) asks: *why is economics not an evolutionary science?* The author analyzes economic systems in the light of the thermodynamic equilibrium of physics. With that, he defends the dynamic aspect of the economy and contests the perception that systems always tend towards equilibrium. For the scholar, economies are in a constant state of evolution. Boschma and Frenken (2006), with a similar motivation, ask: *why is economic geography not an evolutionary science?*

In addition, the authors trace the evolution of EEG as a relevant paradigm in the debate and emphasize its complementarities with other schools.

Initially, Boschma and Frenken (2006) emphasize the theoretical aspects that characterize NEG. As the main exponent of the school, Krugman (1991), according to these authors, expands the neoclassical prescription to explain the processes that determine the agglomeration of activities. He proposes that the agglomeration of production, knowledge, and capabilities emerges as the result of agents' rational decisions. Furthermore, Boschma and Frenken (2006) highlight two main points about Krugman (1991)'s contributions. First, the demonstration that agglomeration can occur without assuming regional differences and external economies is sufficient by assuming the presence of imperfect competition in the market and the existence of increasing returns to scale. Second, Krugman (1991)'s model proved to be extensible in several directions and could explain other phenomena.

The institutional economic geography approach, in turn, cannot necessarily be considered a fully articulated school. Theorists of this tradition argue that differences in institutions determine decision-making and economic behavior. As Boschma and Frenken (2006) argue, institutional differences occur at different levels and may be present between firms or between locations. Finally, these scholars assess that NEG and IEG evolved independently and argue for the evident emergence of an evolutionary approach that communicates with the previous ones but carries idiosyncratic traits.

As demonstrated by the seminal article by Boschma and Lambooy (1999), concepts related to evolutionary economics can be adapted to understand the discussions around economic geography. To understand the object of study of this tradition, Boschma and Frenken (2011, p. 295) define that:

> "Evolutionary Economic Geography (EEG) explains the spatial evolution of firms, industries, networks, cities and regions from elementary processes of the entry, growth, decline and exit of firms, and their locational behavior."

With this quote, it is possible to delve deeper into the conceptualization that Boschma and Frenken (2006) make about EEG in relation to neoclassical and institutional approaches. They define EEG as a third way in the dispute over the understanding of economic progress. That is, the evolutionary approach is based on two fundamental principles: the idea of bounded rationality and the emphasis on routine behavior. In other words, the main objective of EEG is the concern to understand the process of evolution of organizations, which, due to their individual routines built over time, compete with each other. As defined by the influential contribution of Nelson and Winter (1982)[3] in

[3] Nelson and Winter (1982, p. 14) call routines not only the technical procedures of each firm, but also the hiring/dismissal processes, inventory control, investment policies, advertising and business strategies to increase their market power, among other things. They explain that "in our evolutionary theory, these routines play the role that genes play in biological evolutionary theory".

proposing an evolutionary theory for the understanding of economic change, routines are any regular patterns of behavior that characterize a firm. Finally, this allows the EEG to occupy an intermediary position between institutional and neoclassical conceptions of economic geography.

The EEG's intermediate position is supported by the similarities and differences that the approach has in relation to the institutionalist and neoclassical traditions. This process is manifested in three central points. First, the EEG connects with the NEG and distances itself from IEG by understanding the usefulness and fundamentality of formal modeling, while institutionalist authors reject this attribute, arguing that formal models are unable to capture important qualitative factors specific to each location. The second point focuses on the basic principles of each of the theories. EEG and IEG criticize the hypotheses of neoclassical theory, which are based on the modeling of a representative agent that always maximizes its utility. In contrast, evolutionary and institutionalist authors postulate that economic agents are shaped by limited rationality and base their actions, respectively, on their routines and institutions. Finally, there is a point of divergence in evolutionary thought in relation to the other two. The analysis of current issues, according to the EEG, must be understood fundamentally on the basis of the preexisting conditions. That is, there is a critique of the static analysis of NEG and IEG. According to the evolutionary approach, the state of affairs must be understood dynamically, so that it is possible to comprehend current constraints as a result of the state of affairs in previous periods (BOSCHMA; FRENKEN, 2006, p. 280).

Once the differences between schools are understood, it is possible to highlight the main aspects that characterize economic geography as an evolutionary science. In this regard, Boschma and Lambooy (1999) argue that the evolutionary approach holds fundamental concepts to explain some of the main phenomena in the study of economic geography. Concepts such as path dependence, routines, and increasing returns are essential to explain, among other things: the problems of regional adjustments that differentiate the capacity of one locality from another to generate, absorb, or imitate a new technology; the collective learning process located in different regional contexts; and the spatial formation of industries as an evolutionary process and a consequent problem of regional lock-in. Thus, the use of such concepts allows economic geography to be understood from an evolutionary perspective.

In this sense, Boschma and Frenken (2006, p. 291) summarize the key aspects of the evolutionary interpretation of economic geography. Among other things, they claim that EEG combines both inductive and deductive theorizing with formal modeling. Second, the approach takes firms and their routines as the basic unit of analysis. Third, according to the authors, EEG assumes that the success or failure of a firm is a consequence of routines that were built in the past. Fourth, EEG interprets that institutions primarily impact innovations, in a generic sense, but are evolving in tandem with technologies over time and differently across regions. Fifth, EEG explains regional economic development from the structural changes that occur at the levels of firms, sectors, institutions, and different territorial levels. These and other characteristics allow Boschma and Frenken (2006, p. 292) to conclude by saying that:

> "Thus, Evolutionary Economic Geography makes use of formal theorizing grounded in more realistic assumptions (like bounded rationality), but it also conducts case-study approaches that analyse regional specificities from a dynamic perspective. In short, evolutionary scholars favour methodological pluralism."

The spatial evolution of firms, industries, cities, and regions is, therefore, a consequence of structural change over time. As previously defined by Boschma and Frenken (2011), regional development is understood in light of the entry and exit of firms and the evolution and differences between their routines. This theoretical framework supports the analyses that began with reassessments of the contributions of Jacobs (1969) and Marshall (1920) and delved deeper into the process of economic diversification at the regional level. Thus, the next sections will detail the evolution of this literature, leading to the understanding that relatedness is the driving force behind the process of regional diversification.

1.3 Regional Economic Diversification

1.3.1 Specialization versus Diversification

The initial question in this literature was whether a firm primarily benefits from other local firms within the same industry (Marshall externalities) or from local firms in different industries (Jacobs externalities). Glaeser et al. (1992) addressed this debate while analyzing the growth of large industries in US cities. They sought to determine the main source of spillovers that encouraged industrial growth: industry specialization within a region or the presence of a diverse range of industries. Ultimately, the authors found that job growth in the industries studied was more closely associated with local competition and urban diversification rather than regional specialization. Consequently, it was concluded that knowledge spillovers occur more significantly between industries than within industries. This evidence supports the notion that diversification is a key driver of growth, aligning with Jacobs (1969)'s perspective.

Other scholars, evaluating the source of these spillovers under the same principles, have arrived at slightly different results. Henderson, Kuncoro and Turner (1995) assess the evolution of employment in industries in different cities from 1970 to 1987. They find that the main source of externalities depends on the maturity level of the industry. Among younger, and especially high-tech, industries, externalities resulting from industry diversification in the region (Jacobs) proved to be the most influential. On the other hand, among the more mature industries, the greatest influence is given by regional specialization, thus highlighting the locational externalities (Marshall).

Boschma and Frenken (2011, p. 299) better contextualize the distinction between Marshall and Jacobs externalities. Agglomeration economies arising from regional specialization (Marshall) come from dense and specialized labor markets, access to specialized suppliers, and knowledge spillovers. On the other hand, to exemplify diversification externalities, they argue that Jane Jacobs was a pioneer in recognizing how the deep division of labor in cities could benefit urban economic growth by also stimulating opportunities for innovation, not just for reasons of efficiency. In short, Jacobs' externalities are supported by the idea that diversified cities induce spillovers between industries and recombinant innovations.

These questions were also addressed in empirical analyses for Brazil. Simoes and Freitas (2014) use formulations made by Glaeser et al. (1992) to identify the impact of external economies on the productivity of industries in Brazilian regions. They found evidence that productivity benefited both from specialization (Marshall) and diversification (Jacobs) externalities between 2000 and 2010. Although Marshall externalities were significant in all technology sectors, the greatest impact of Jacobs externalities was present in medium-high and high technology segments.

However, there are still relevant discussions that go beyond this dichotomy. To demonstrate this, Boschma and Frenken (2011) list some movements in the literature that aim to assess the nature of agglomeration economies, but that are not restricted to the idea of evaluating the contrast between Marshall and Jacobs' externalities. The authors emphasize that there are studies concerned with deepening the investigation of transmission channels through which knowledge spillovers occur. Additionally, they also highlight the growing volume of analyses on agglomeration economies that focus on the firm level. From an evolutionary perspective, as firms are very heterogeneous and consist of different routines and capabilities, externalities should influence them in different ways. Finally, and most importantly for us, Boschma and Frenken (2011) emphasize that, beyond the aforementioned dichotomy, knowledge spillovers occur when there is cognitive proximity, neither too close nor too far, between economic agents. This perception is common to several articles that focus on evaluating the process of economic diversification, highlighting the different ways in which it occurs regionally.

1.3.2 Relatedness as a Driver of Regional Diversification

Diversification alone is not enough to generate spillovers. Frenken, Oort and Verburg (2007) argue that a favorable environment for knowledge spillovers to occur is present in economies characterized by productive structures filled with sectors that share similar capabilities. In other words, more than just having variety, as formulated by Jacobs (1969) some level of relatedness between sectors/firms in the economy is also necessary. However, in addition to this type of diversification, there is also the possibility of unrelated variety, where the productive structure is better suited to absorb shocks in specific sectors without impacting the entire economy.

Frenken, Oort and Verburg (2007) discuss three points of the relationship between diversification and economic development. First, they emphasize the fundamental connection between variety and growth, as previously presented (JACOBS, 1969; GLAESER et al., 1992). Second, they specifically highlight the link between variety and unemployment, based on the interpretation that portfolio protection exists. In a nutshell, a region that competitively produces a wide variety of sectors has an economy that is less susceptible to the negative effects of demand shocks, as it adopts a risk-spreading strategy. On the other hand, regional specialization increases the vulnerability to absorb shocks, resulting in high unemployment rates and slowing growth. Finally, they argue that economies with lower levels of diversification tend to suffer from the evolution of structural unemployment and consequent stagnation. By understanding the various ways in which the diversification of productive structures can impact regional economic development, the objective becomes to discuss which variety contributes more to growth or stability - related or unrelated variety.

To evaluate the two types of variety, Frenken, Oort and Verburg (2007) analyzed two hypotheses using employment data in the Netherlands from 1996 to 2002: (i) related diversification is associated with Jacobs externalities, resulting from spillovers, and contributes to employment growth; and (ii) unrelated diversification is associated with a portfolio strategy that avoids regional increases in unemployment. They found that related variety contributes to employment growth, while unrelated variety is negatively linked to unemployment growth. These findings validate, respectively, that Jacobs externalities are major drivers of regional economic growth and that the presence of unrelated sectors acts as a defense against unemployment shocks. Finally, they stress that a regional diversification policy must combine the benefits of production in related sectors and be complemented by national policies focused on more generic technologies.

Boschma and Iammarino (2009), in turn, examine the impact of external trade linkages on regional development. Their argument is that the mere influx of knowledge flows is not enough to ensure economic growth. It is crucial that firms possess the absorptive capacity to effectively translate this knowledge into regional growth. Boschma and Iammarino (2009) find that commercial similarity between regions does not have a significant effect on regional growth. Surprisingly, they discover that regions benefit more from extraregional knowledge linkages when these connections originate from non-similar but related sectors. In other words, regions that are linked to industries that share related capabilities and skills experience more positive effects on their growth, emphasizing the importance of relatedness in fostering regional economic development.

With the methodological and empirical advancements introduced by Hidalgo et al. (2007), the concept of related diversification has become more comprehensible and accessible. The introduction of the Product Space, which establishes connections between products based on their co-occurrence in countries' export portfolios, has clarified various issues. One of the significant findings is the structural differentiation between developed and underdeveloped countries. Hidalgo et al. (2007) demonstrate that industrialized countries competitively produce goods in sectors located in the core of the Product Space, densely connected and represented by industries such as chemicals, machinery, and equipment. In contrast, Latin American and Caribbean countries are positioned on the periphery of the Product Space, characterized by less dense connections and dominated by labor and land-intensive sectors. Moreover, the related diversification process highlights that the path to the richest parts of the Product Space is considerably more challenging for regions located in the peripheral area, as the connectivity between sectors is less dense in those regions.

The progress made in measuring concepts such as proximity between sectors has led to a better empirical understanding of the spatial concentration of skills, knowledge, technology, and capabilities. This unprecedented measurement has facilitated the evolution of empirical analyses on the determinants of the location of productive activities. Several studies have utilized these new concepts to establish a strong relationship between the probability of a sector being competitively produced in a region and its proximity to other sectors in the local portfolio. These contributions were consolidated by Hidalgo et al. (2018) around what they called "The Principle of Relatedness", a phenomenon that has been studied for many years but could only be recently formalized empirically.

The Principle of Relatedness has been tested in analyses at different spatial levels (national and regional) and using various strategies. One notable study conducted by Hausmann and Klinger (2007) focused on analyzing the Product Space at the international level. By calculating a density indicator to measure the proximity between products and the structures of countries, the authors drew significant conclusions. The evidence demonstrated that density is a crucial determinant of countries' structural change. In other words, when there are changes in the pattern of specialization, countries tend to shift mainly towards nearby products. However, this analysis also raises important questions that require further exploration in the literature. For instance, what are the factors influencing structural transformation? What determines the possibility of countries moving to more distant sectors of the Product Space? Additionally, there is a need to understand the potential of public policies in promoting and facilitating such movements.

At the regional level, the results are also robust and encompass regions from different countries. Neffke, Henning and Boschma (2011) employ an indicator based on the co-occurrence of products in Swedish factories to measure proximity between sectors. They argue that structural change must follow an evolutionary logic of regional branching, emphasizing an aspect of path dependency. The results they found reinforce the notion that new economic activities are not randomly developed in any sector but rather emerge in sectors that are technologically related to the existing sectors within the local portfolio.

However, it is important to note that the literature on complexity has not been included in the analyses conducted here, and it represents a significant path for future research. This is because sectoral and regional complexity are likely to impact the process of structural transformation and could provide valuable insights into the dynamics of economic development and diversification.

Kogler, Rigby and Tucker (2013) conducted a simpler analysis using US patent data, where they constructed a Knowledge Space similar to the Product Space. They measured the "technological distance" between patents and found that higher levels of relatedness indicate a clustering of technological classes that are closely located to each other in the Knowledge Space.

Similarly, Rigby (2015) used US patent data and econometric tests to reach similar conclusions, validating the hypothesis that technological diversification is influenced by the existing knowledge structure in a city. Both studies raise questions about how cities can venture into more unexplored areas of the Knowledge Space, underscoring the significance of the literature that investigates diversification policies and strategies.

Boschma, Balland and Kogler (2015) conducted an analysis similar to Rigby (2015) but with the addition of considering structural characteristics of the technological classes and cities in their estimations. Despite this modification, the results remained consistent with previous findings. They observed a strong and positive relationship between the proximity of technology to the existing knowledge structure and its emergence in the local portfolio. However, similar to Neffke, Henning and Boschma (2011), complexity was not included in these tests, representing an important factor that requires further examination to better characterize the diversification process.

In the context of Brazil, Freitas (2019) conducted an evaluation of the entry, exit, and maintenance of productive activities in regional portfolios. The findings confirmed the hypothesis that density is positively correlated with the probability of entry and negatively correlated with the probability of exit of productive activities. Furthermore, by incorporating sectoral complexity into the analysis, it was concluded that regions are less likely to develop specializations in complex productive activities. However, the study did not fully explore the potential heterogeneity in this relationship across regions with different levels of complexity. In other words, regional complexity also influences the manner in which sectoral complexity impacts the overall diversification process in different regions.

Francoso, Boschma and Vonortas (2022) performed a similar analysis considering employment and patent data for Brazil. They aimed to assess the probability of entry of new activities or new technological classes into local portfolios. The results found show the conclusions already mentioned by Freitas (2019). Whether for productive activities or for technological classes, density and complexity are, respectively, positively and negatively linked with entry probability. However, the differentiation of regions by complexity is also little explored here, and the analysis is not complete due to the absence of the effects on the probability of leaving.

Finally, Pinheiro et al. (2022) delve deeper into the influence of complexity on the process of economic diversification in European regions. However, their findings reveal a concerning diversification pattern when contrasting economically more advanced and backward regions. Advanced regions tend to expand their specialization in more complex related sectors, while backward regions focus on less complex related sectors. This pattern of diversification leads to a divergent and polarized development between regions, exacerbating inequalities. In the following sections, we will focus on this dynamic in two ways. Firstly, we will review the existing literature on complexity and inequality, highlighting their limitations. Secondly, we will discuss the literature that proposes diversification policies, which are crucial in mitigating regional disparities and breaking the feedback loop of inequality that keeps regions trapped in their current state.

1.3.3 Relatedness as Driver of Regional Inequality

The literature on complexity and inequality is relatively recent, leaving several questions still unanswered. A seminal contribution by Hartmann et al. (2017a) was the first to establish a relationship between the Economic Complexity Indicator (ECI) and income inequality in countries, measured by the Gini index. The authors found a negative correlation between these indicators, indicating that economic complexity is a negative predictor of inequality in a country. Subsequent analyses have followed a similar approach.

Lee and Vu (2020) challenged the initial results of Hartmann et al. (2017a). They employed an alternative econometric specification to analyze the relationship between complexity and inequality at the country level, and their findings revealed a positive correlation between these dimensions.

A similar outcome was also observed in the study conducted by Chu and Hoang (2020) who further tested this relationship by considering an alternative complexity indicator that incorporates the difficulty of producing goods. Remarkably, the results remained consistent with a positive association between complexity and inequality.

In contrast, Sepehrdoust, Tartar and Gholizadeh (2022) focused on evaluating middle-income countries and concluded that the relationship between complexity and inequality becomes negative beyond a certain threshold.

Therefore, the discussion of complexity and inequalities primarily centers on income inequality and analyzes at the national level. However, the regional aspect in this approach has been enriched by the insights presented by Hartmann and Pinheiro (2022). The authors make two significant contributions to the objectives of this work. First, they reassess the connection between complexity and income inequality while considering different levels of spatial aggregation. Consequently, they conclude that the relationship is inverted at regional levels when compared to the national level. The second contribution addresses a cause of this phenomenon. Hartmann and Pinheiro (2022) argue that analyzing this relationship at highly aggregated levels conceals important local dynamics that impact economic development. In other words, complex activities are challenging to develop and tend to concentrate geographically, exacerbating inequality between regions. Thus, regional inequality becomes a crucial aspect of the analysis.

Hartmann and Pinheiro (2022) reveal that assessments at the regional level exhibit an inverse link compared to the findings of Hartmann et al. (2017a) where complexity is a negative predictor of income inequality. This inverse relationship is also supported by earlier analyses that identified a positive connection between innovation and inequality (LEE, 2011; LEE; RODRIGUEZ-POSE, 2013). More recent works employing complexity indicators have also found a positive association with wage and income inequality. For instance, Sbardella, Pugliese and Pietronero (2017) combined regions' rankings in the Fitness complexity indicator and GDP per capita, observing a positive effect on wage inequality in US counties. In a different context, Marco, Llano and Pérez-Balsalobre (2022) demonstrated similar dynamics in Spain, highlighting a trade-off between complexity, income inequality, and environmental quality.

In Brazil, the analysis that explores the connection between complexity and income inequality is conducted by Morais, Swart and Jordaan (2021). They examine this relationship at the level of Brazilian states and surprisingly find a negative association. However, it is worth noting that most states have dimensions similar to countries, which can make it challenging to capture the effects that could determine a positive relationship between complexity and inequality at the regional level. To better understand the dynamics, researchers have mentioned smaller aggregations, such as meso-regions, which reveal signs of an inversion of the relationship, as demonstrated by Hartmann and Pinheiro (2022).

To understand this dynamic, Hartmann and Pinheiro (2022) present three main reasons that may explain the paradox between national and regional analyses. Firstly, they argue that higher levels of complexity indicate positive institutional development. In other words, competitive production in more complex sectors requires modern and inclusive institutions that enable economic agents to learn from each other effectively. Secondly, more complex countries tend to outsource production to less sophisticated sectors that employ lower-skilled and lower-wage labor. Thirdly, the authors emphasize the influence of agglomeration effects at the regional level, which conditions the analysis differently. The concentration of more complex activities in large centers and the coexistence of both more and less sophisticated activities in regions prevent similar outcomes to national-level analyses (BALLAND et al., 2020). In this context, spatial agglomeration effects such as migration, labor market disparities, and the spatial concentration of sophisticated activities contribute to perpetuating polarized center-periphery productive structures and regional inequality (HARTMANN; PINHEIRO, 2022; FURTADO, 1964; PREBISCH, 1950).

As highlighted by Hartmann and Pinheiro (2022), this ambiguous effect is also evident in the relationship between diversification and human development (LAPATINAS, 2016). On one hand, economic diversification can promote the democratization of access to basic services, such as health and education, and expand opportunities for individuals. However, on the other hand, it perpetuates a process of regional economic polarization, leading to winners and losers.

Balland et al. (2020) also shed light on this issue, demonstrating that more complex activities tend to concentrate in large centers. In other words, the level of complexity of a sector influences its spatial clustering. As a result, the competitive production of new complex sectors is more likely to occur in these large centers due to their greater absorptive capacity. This process, in turn, repels less desirable and less related industries away from these centers, leading to divergent economic evolutions between regions or cities and exacerbating intra- and inter-local inequalities.

The tendency to diversify into related industries, as summarized by the Principle of Relatedness, leads to vastly different paths for less and more complex regions. Less complex regions tend to intensify their production in more related sectors, which often rely on less qualified labor and are based on natural resources. In contrast, more complex regions focus on activities that are more knowledge-intensive and inclusive, as these sectors are more closely related to complex productive structures (HARTMANN; PINHEIRO, 2022). As emphasized by Hartmann et al. (2020), while the composition of the productive structure is not the sole factor causing inequality, it is a crucial structural characteristic that significantly influences available employment and income distribution alternatives. This dynamic underscores that relatedness is a strong driver of regional inequality.

In Brazil, Freitas and Paiva (2015) identified a concerning trend towards the concentration of diversity and sophistication of productive sectors in historically more developed regions. This complexity-based approach revealed that regional economic polarization worsened in Brazil from 2002 to 2014, leading to disparities in economic development.

Similarly, Cimini et al. (2017) focused on Minas Gerais and concluded that the region remains trapped in a vicious cycle, unable to diversify into more complex sectors due to its low complexity. In addition, while studies by Freitas (2019) and Francoso, Boschma and Vonortas (2022) highlighted the influence of relatedness in the diversification process, the role of sectoral and regional complexity has not been explored. As is the case in the general literature, analyses for Brazil would benefit from further research that better characterizes the process of related diversification in the context of complexity and assesses its determinants and consequences.

The analysis of relatedness as a driver of regional inequality in Brazil is ilustrative for several reasons. First, this continuous process of divergence of regional economies has already been detected in other countries, leading to detrimental effects on economic and social progress (IAMMARINO; RODRIGUEZ-POSE; STORPER, 2019). Second, it is crucial to expand geographic wisdom while assessing the regional diversification process, meaning that including regional context specifications in the analysis is essential (BOSCHMA, 2017). Third, Brazil is a country characterized by a high degree of inequality and composed of very different regions, which adds complexity and challenges to the study (MORAIS; SWART; JORDAAN, 2021; FREITAS; PAIVA, 2015). Fourth, there is evidence suggesting that Latin American countries, including Brazil, have intensified their dependence on activities associated with high levels of inequality over the years (HARTMANN et al., 2017b). In this sense, understanding the restrictions that this process imposes on different regions is vital for developing effective regional development policies.

Although the link between complexity and inequalities is a phenomenon that deserves significant attention, the literature on the subject remains limited. Firstly, the existing studies primarily focus on examining the relationship between conventional indicators of complexity and inequality, leaving potential other factors unexplored. Secondly, the analysis of inequality in relation to complexity often remains confined to income distribution indicators, neglecting other forms of inequalities, such as wage disparities between sectors. Thirdly, there is a lack of research assessing how varying levels of complexity influence the driving forces in local economies, such as job creation. In summary, despite the evident perception of divergent development between regions in the literature, only a few studies are dedicated to analyzing this dynamic in-depth.

Pinheiro et al. (2022) and Hartmann and Pinheiro (2022) discuss the inherent unequal nature of related diversification but do not quantify this relationship and its implications. Therefore, the literature still requires answers to pertinent questions. What is the impact of a sector's complexity on the likelihood of its incorporation into local production? How does the magnitude of this effect vary with the complexity of existing structures? These questions should guide our understanding of how inequality is established in the process of diversification between less and more complex regions. Additionally, what are the consequences of diversifying into more complex sectors? How does greater regional complexity influence local job creation, for instance? The answers to these questions are crucial contributions that the literature demands to comprehend the implications of related diversification. Lastly, how challenging is it to break free from this "winners and losers" dynamic? What effective policies can remove regions from a low complexity trap? The literature that discusses practical solutions and policy guidance is the subject of the next section.

1.3.4 Diversification-oriented Policies

The advancement of the literature has primarily led to the formulation of robust methodologies to guide diversification policies. The ability to measure abstract concepts concerning sectors or the productive structure, such as proximity, distance, and density, has enabled the development of practical recommendations for public policies, streamlining the decision-making process for policymakers. Additionally, diversification policies are especially valuable for countries and regions facing stagnation, struggling to diversify into more complex sectors. The extensive literature on diversification methodologies and policy suggestions encompasses regions from various countries.

Hausmann and Chauvin (2015) conducted a study focused on Rwanda, where they explored various strategies considering the country's specific constraints. The study comprised two main approaches: one concentrating on export opportunities in global markets and the other on the regional market. Their methodology involved identifying products that align with Rwanda's current capabilities and possess the potential to expand the country's productive knowledge. The authors emphasized two types of products: less complex ones with low transport costs, suitable for distant markets, and more complex ones with higher transport costs but highly demanded by neighboring countries. Through their analysis, they identified over a hundred new products that are less resource-intensive, positioned at the edge of Rwanda's productive knowledge, and feasible in terms of transport costs.

Hausmann, Santos and Obach (2017) extended their methodology to Panama, proposing a Diversification Opportunity Score. This score is calculated for each sector based on three key dimensions. Firstly, it considers Panama's existing capabilities, taking into account factors such as the value, intensity, and growth of exports of the product. Secondly, it incorporates the current market opportunities, analyzing the value and intensity of imports from various regions. Finally, complexity criteria are evaluated, including industry complexity, density, and the potential gain of opportunity within the sector. This comprehensive score calculation method has been employed in various studies to suggest promising sectors for regional economic development.

In their study, Balland et al. (2018) took a unique approach by presenting a methodological framework for smart specialization policies in European regions. Their framework incorporates measures of both relatedness and complexity to assess potential risks and rewards. The authors propose that sectors with higher complexity offer greater benefits from the policy, while those with lower complexity entail fewer benefits. Similarly, sectors with higher relatedness pose smaller risks, while those with lower relatedness present greater risks.

Furthermore, in their analysis of Paraguay, Hartmann, Bezerra and Pinheiro (2019) present an innovative approach to identifying intelligent diversification strategies. They consider two key dimensions: feasibility and desirability. Feasibility is evaluated through indicators like revealed comparative advantage (RCA) and relatedness. On the other hand, desirability is assessed based on the expected levels of income, complexity, technology, and income inequality associated with each sector. By comparing four different strategies, ranging from conservative, relatedness-based approaches to more ambitious ones, the authors find that setting minimum standards of viability and desirability for new products leads to the most successful policies. This framework helps in identifying and prioritizing sectors with high potential for economic development and greater long-term benefits for the region.

Approaches similar to those mentioned have been developed for Brazil as well. Romero and Silveira (2019) adapted the methodology of Hausmann, Santos and Obach (2017) to create a diversification opportunity score and identify promising sectors for the development of Brazilian states. A similar analysis was conducted by Queiroz, Romero and Freitas (2023), using a scoring framework to propose three promising sectors for diversification in each Brazilian state and simulate the impact on local productive structures in terms of employment. These analyses offer valuable insights and enable the suggestion of more targeted and effective diversification policies at even more granular levels. Additionally, Romero et al. (2022) adapted the scoring method and employed principal component analysis (PCA) to determine the weights of each dimension in the scoring indicator. By using this approach, they identified promising economic activities for the development of the city of Belo Horizonte.

Finally, a common aspect present in all the literature proposing methodologies and policies for guiding regional diversification is the emphasis on relatedness as a fundamental aspect of these strategies. The proximity of new products to existing local structures serves as a crucial criterion for identifying the appropriate path to follow.

However, as Boschma (2021) summarized, a strategy focused solely on related diversification faces four primary challenges. First, it tends to be comparatively conservative, offering fewer incentives to push the technological frontier and foster innovations (BOSCHMA; LAMBOOY, 1999). Second, related diversification can occur naturally, making specific policies focusing on related variety potentially unnecessary in some cases. Third, if diversification is too closely related, it may limit regions from exploring other opportunities for greater diversification, leading to a state of lock-in. (BOSCHMA; LAMBOOY, 1999; BOSCHMA, 2005; BOSCHMA; IAMMARINO, 2009; BALLAND; BOSCHMA, 2021). Lastly, as mentioned in the previous section, the focus on related diversification may primarily benefit regions with high complexity, leaving lagging regions predominantly producing low-complexity sectors. In other words, this process can perpetuate a loop that intensifies regional inequality (PINHEIRO et al., 2022). To address these challenges, comprehensive policies are essential, striking a balance between related and unrelated diversification and considering the unique characteristics and complexities of each region.

The alternative to freeing regions from a low-complexity trap state lies in these policies. At the macro level, there is substantial evidence of countries that have diversified into unrelated sectors and reaped future economic benefits. Pinheiro et al. (2018) examined the economic diversification trajectory of 93 countries between 1970 and 2010 and found that only 7.2% diversified into unrelated products. However, their analysis also revealed that in the short term, these countries experienced a higher economic growth of 0.5% compared to other countries with similar characteristics. While this macro-level evidence is significant, it also emphasizes the need for further research to assess this dynamic at the regional level. The presence of divergent and polarized growth between regions underscores the necessity for diversification policies capable of breaking this pattern and curbing the evolution of regional inequalities. Tailored strategies and targeted interventions at the regional level can contribute to promoting more inclusive and balanced economic growth across regions.

1.4 Concluding remarks

This chapter sought to discuss the literature on economic diversification, complexity, and inequalities from a regional perspective. The literature considered here is based on the interpretation that economic geography is an evolutionary science. This implies that regional economic development is primarily a consequence of structural changes that occur over time at the level of firms, sectors, institutions, and various territorial levels (BOSCHMA; FRENKEN, 2006). This interpretation provides essential concepts to comprehend the dynamics of regional economic disparities through the economic complexity approach (HIDALGO et al., 2007; HIDALGO; HAUSMANN, 2009). Notably, concepts like path dependence, routines, and increasing returns to scale are central to understanding local economic development and its variations (BOSCHMA; IAMMARINO, 2009).

The recent resurgence of the discussion present in the works of Marshall (1920) and Jacobs (1969) shed new light on the role that agglomeration economies assume in the context of regional economic development. The issue of whether spillovers are more recurrent from specialization (Marshall) or diversification (Jacobs) externalities culminates in the seminal contribution by Glaeser et al. (1992). Results found indicate a predominance of beneficial effects of diversification on employment growth, highlighting the Jacobs externalities. From this important contribution, an extensive literature has studied the existing differences in the dynamics of diversification and its importance in regional development.

Frenken, Oort and Verburg (2007) delved deeper into the impact of diversification on regional development, considering the difference between related and unrelated diversification. The evidence found reaffirmed the authors' two main hypotheses. First, when diversification moves into activities that are more closely related, in terms of similar capacities to be produced, there is a positive effect on regional employment growth. On the other hand, when diversification is directed towards unrelated sectors, an economy portfolio strategy is established that protects it from possible sectoral crises. From there, several studies are derived that highlight relatedness as a driver of diversification and regional economic growth.

However, relatedness has a central consequence for regional economic development. The tendency of economies to diversify into sectors with more related capabilities creates a logic of "winners and losers", in which the most advanced regions have the opportunity to diversify into highly complex activities, while the most backward are restricted to less complex related activities (PINHEIRO et al., 2022). This process creates a feedback loop that continually aggravates inequalities between regions. Furthermore, Balland et al. (2020) pointed out that complex activities have been continuously concentrated in large cities, demonstrating once again that the growth of spatial inequality may be associated with the increase in the complexity of economies. In this context, the literature that discusses and proposes methodologies to guide diversification policies assumes a central role. However, the related diversification as a driver of regional inequality is still little quantitatively characterized, and the literature is scarce in assessing its determinants and consequences.

The main contribution of this chapter is to point out issues that still need to be addressed in the literature. Initial contributions studied only the influence of related diversification on the regional economic growth process (FRENKEN; OORT; VERBURG, 2007), without focusing on alternative consequences of this process. The perception that relatedness drives the worsening of regional inequalities is discussed in the literature, but there is a lack of studies that characterize it and quantify its consequences (HARTMANN; PINHEIRO, 2022; PINHEIRO et al., 2022). Questions about the influence of sectoral and regional complexity on the diversification process are central to understanding this economic divergence. These gaps set the stage for the analyses that will be made in Chapters 2 and 3.

2 Why is the news good for some and bad for others?

2.1 Introduction

Hausmann and Klinger (2007) and Hidalgo et al. (2007) showed 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; OORT; VERBURG, 2007), Italian (BOSCHMA; IAMMARINO, 2009), Swedish (NEFFKE; HENNING; BOSCHMA, 2011), American (BOSCHMA; BALLAND; KOGLER, 2015; ESSLETZBICHLER, 2015), and Brazilian regions (FREITAS, 2019; FRANCOSO; BOSCHMA; VONORTAS, 2022) follow a similar trajectory. In other words, regional economic diversification tends to prioritize sectors that are already related to the existing productive structure.

The findings from studies by Freitas (2019) and Francoso, Boschma and Vonortas (2022) underscore the importance of relatedness in the regional diversification process in Brazil. However, these findings also reveal 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.

This chapter 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 Francoso, Boschma and Vonortas (2022) focused solely on the influence on the probability of entry, the primary contribution of this chapter is to investigate the effect of complexity on diversification in regions with distinct complexity levels.

To achieve this objective, this study utilizes formal employment data in productive activities across micro-regions in Brazil from 2006 to 2021. 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 holds true, as in Freitas (2019) and Francoso, Boschma and Vonortas (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) into: Low, Medium-Low, Medium-High and High.

These complexity-based regional groups enable the examination of novel 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.

The chapter is organized as follows: Section 2 discusses the literature on the subject, bringing relevant and similar contributions applied to different regions across the world and which support the work developed here. Section 3 demonstrates the calculated indicators, the description of the database and the econometric specifications used. Section 4 is a descriptive analysis of the data, exposing the results of the indicators, a brief description of Brazil's regional diversification over time and a demonstration of the complexity groups. Section 5 brings the results of the econometric tests and section 6 ends with the final considerations.

2.2 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; OORT; VERBURG, 2007; NEFFKE; HENNING; BOSCHMA, 2011; BOSCHMA; BALLAND; KOGLER, 2015; FRANCOSO; BOSCHMA; VONORTAS, 2022). Moreover, the integration of these concepts into economic geography builds upon seminal studies that have examined regional growth and its driving factors (GLAESER et al., 1992; HENDERSON; KUNCORO; TURNER, 1995). Therefore, this section aims to discuss the evolution of empirical literature that evaluates regional economic growth and development, with a focus on describing and relating how these studies explain the related diversification process.

Glaeser et al. (1992) argue that the growth of cities is directly influenced by competition and diversity within industries. The authors present two key exercises that support this conclusion. Firstly, they examine whether industrial concentration or competition impact employment growth in city-industry pairs between 1956 and 1987. Based on their findings, they conclude that cities with higher concentration in a single industry experience slower employment growth, whereas cities with a greater number of firms per worker than the national average exhibit stronger growth. Secondly, they compare employment growth in the top four industries with growth in industries outside of the top four. This analysis leads them to another conclusion: diversification also contributes to employment growth. These results have had a significant impact on regional literature, as they provide evidence for the importance of Jacobs (1969) externalities in regional economic development.

The findings of Glaeser et al. (1992) were further enhanced by the evidence presented by Henderson, Kuncoro and Turner (1995). These authors investigate the influence of diversification and concentration externalities by taking into account industry type and maturity level. Initially, using a tobit model, they determine that a higher concentration of employment in a specific industry in the past has a more significant impact on current employment growth. On the other hand, the authors employ a probit model to identify the determinants for attracting high-tech industries. They compare various independent variables, including diversity and historical concentration of employment in the industry. The results indicate that concentration externalities have a minor role, while diversification externalities are crucial for attracting high-tech industries. In summary, the scholars conclude that diversification is important for attracting new industries, while concentration plays a key role in retaining them.

In addition to the aforementioned discussions, several authors have delved deeper into the study of diversification to better understand its role in local development. Frenken, Oort and Verburg (2007) have made efforts to assess the impact of diversification on employment and unemployment growth in Dutch regions, examining two dimensions: related and unrelated diversification. The former is measured using the weighted sum of entropy within each two-digit sector, while the latter focuses on entropy at the two-digit level. The authors employ ordinary least squares (OLS) estimations and test the robustness of their findings by using different time periods. Additionally, considering the possibility of spatial dependence among the independent variables in the model, the scholars also conduct estimations using an average window approach, which assigns weights to the values of the region and its neighboring regions. Ultimately, all the models provide support for the initial hypotheses: related diversification has a positive influence on employment growth, while unrelated diversification has a negative influence on unemployment growth.

In a complementary study, Boschma and Iammarino (2009) conduct a similar analysis to assess the regional economic growth of 103 Italian provinces. The authors examine the effects of variety (without specifying the type), related variety, and unrelated variety on employment growth, value added, and labor productivity across regions. Similar to Frenken, Oort and Verburg (2007), variety measures are calculated using the entropy indicator. The estimations are performed using OLS with fixed effects for macro-regions. The results indicate that variety alone does not have a significant impact on the economic growth of the provinces, thus highlighting the importance of differentiating between types of diversification. However, when specifically specified, related variety demonstrates robust effects on regional growth, while unrelated variety does not show statistical significance.

These results have paved the way for a series of subsequent articles that have led to the synthesis of the empirical Principle of Relatedness. This principle describes the likelihood of a region entering or exiting productive economic activities based on the presence of related activities in that location. Researchers have utilized various sources of information to calculate relatedness, also referred to as proximity or density, between activities and their respective local productive structures, considering that the capacities involved in the production process may not be perfectly observed. For instance, co-exporting of products, co-occupation in productive activities, and input-output flows between industries have been employed as indicators. The utilization of diverse sources has enabled different authors to establish a robust relationship between the probability of a region specializing in a particular activity and the existing related capabilities in that region (HIDALGO et al., 2018, p. 452). This principle, therefore, highlights the significance of related diversification in the development process of regions and countries and has provided support for the articles referenced in this work.

The significance of related diversification is empirically demonstrated at the macro level in the work of Hausmann and Klinger (2007). Using foreign trade data, the authors examine the impact of density[1] on a country's ability to competitively enter a specific sector. Through estimations using OLS with clustered standard errors, they find that a smaller distance between a country's productive structure and a specific product significantly increases the probability of the country moving towards that sector. Specifically, a one-standard-deviation increase in density, on average, raises the indicator of revealed comparative advantage in that sector by 0.366. The estimates are robustness-checked using different time periods, unscaled density indicators, and alternative definitions of "export". However, it is within the regional literature that a more detailed exploration of the diversification process and the importance of product proximity to the productive structure for driving structural changes has been undertaken.

[1] It is a "measure of how close is a country to each of the products it currently does not export with comparative advantage" (HAUSMANN; KLINGER, 2007, p. 19-20).

The work of Neffke, Henning and Boschma (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 (i) the probability of entry, (ii) maintenance, and (iii) 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 OLS, probit, and logit models, with the authors controlling for employment information in both the region and industry. 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. This analysis has been widely replicated by other researchers to explore additional influences and locations.

A similar analysis was conducted by Essletzbichler (2015) using data from 360 US metropolitan areas. The approach followed the same logic as Neffke, Henning and Boschma (2011), with the exception that only logistic regressions were employed for the estimations. 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, Henning and Boschma (2011) determined relatedness based on the occurrence of products from distinct industries in manufacturing plant portfolios, Essletzbichler (2015) measured relatedness by analyzing the intensity of flows between pairs of industries using input-output relations. The author utilized input-output flows between 563 US industries. 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 (2015) 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, while controlling for a time-lagged measure of knowledge flow directed to the city. 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 exit by 2.72%. Other researchers have also employed patent data for similar analyses.

Boschma, Balland and Kogler (2015) conducted a similar analysis to Rigby (2015) 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 between 1976 and 2010, 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 measures of relatedness were tested, outliers were excluded, and 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. This finding highlights the crucial role of relatedness as a driving force for technological change in the examined US cities (BOSCHMA; BALLAND; KOGLER, 2015, p. 244).

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 assumption was that topics that frequently co-occur in the same article are related and possess similar capabilities. To estimate the effect of lagged relatedness on the entry and exit of topics in the cities' portfolios, the authors employed OLS models. The model specification resembled that of Neffke, Henning and Boschma (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 suggests that scientific relatedness plays a significant role in the dynamics of biotech research topics in cities.

This method was also used to assess regional diversification in Brazil. Freitas (2019) utilized employment data from Brazilian micro-regions 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.

Francoso, Boschma and Vonortas (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, 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 Francoso, Boschma and Vonortas (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). However, this structural characteristic of the Brazilian regional diversification process was not the primary focus of the analyses conducted by Freitas (2019) and Francoso, Boschma and Vonortas (2022). The former examined only the diversification of the most complex micro-regions, 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 Francoso, Boschma and Vonortas (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. The differences between regions play a significant role in shaping their diversification trajectories and can contribute to divergent development outcomes. 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 the Brazilian context. 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, two hypotheses will be tested in this chapter:

- *Hypothesis 1*: 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.

- *Hypothesis 2*: Among less complex regions, greater sector complexity decreases the probability of sector entry, and simultaneously acts as a significant factor driving sector exit.

In this case, Hypotheses 1 and 2 have not been tested together in previous studies and represent the main contribution of this chapter. Francoso, Boschma and Vonortas (2022) partially examined Hypotheses 1 and 2, 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 confirm Hypotheses 1 and 2. In addition, the authors use a simple division of regions, more aggregated, since this it is not the main objective of the article. On the other hand, Freitas (2019) only tested Hypothesis 1 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 2 as it did not include less complex regions.

2.3 Data and Method

This section outlines the procedures for calculating the complexity and relatedness indicators. Additionally, it provides details regarding the database used for calculating the indicators, the level of regional and activity aggregation chosen, and the time period considered for the analysis. Furthermore, the model specification is presented, including the variables and the methods employed. Although briefly discussed, the chosen methods are explained to provide an overview of the analytical approach.

2.3.1 Complexity Measures

The methodological innovations introduced by Hidalgo et al. (2007) and Hidalgo and Hausmann (2009) have provided a fresh perspective on the literature of economic development. These authors argue that variations in productivity among countries can be explained by their unique sets of capabilities and the interactions among them. The inability to import specific individual skills results in a differentiation of the available "tools" for production in each country. Countries with a greater diversity of capabilities are more likely to produce sophisticated goods with higher economic value. As a result, the introduction of indicators capable of measuring these capabilities has opened up new possibilities for in-depth assessments of the influence of productive structures on growth differentials among countries.

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, whether in terms of exports, production, or employment, depending on the specific data source. 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:

$$RCA_{i,j} = \frac{X_{i,j}}{\sum_i X_{i,j}} \div \frac{\sum_j X_{i,j}}{\sum_j \sum_i X_{i,j}} \tag{2.1}$$

Where $X_{i,j}$ 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:

$$D_j = k_{c,0} = \sum_i M_{i,j} \tag{2.2}$$

$$U_i = k_{i,0} = \sum_j M_{i,j} \tag{2.3}$$

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_i). In this framework, complexity is measured based on both diversification, which refers to the production of a wide variety of goods and services by countries, and ubiquity, which captures goods that require specialized knowledge and are produced in a limited number of locations where the necessary skills exist. Hence, a complex country or product is characterized by high diversification and low ubiquity. The binary matrix ($M_{i,j}$) 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. The iterated combinations are formally presented below:

$$k_{j,N} = \left(\frac{1}{k_{j,0}}\right) \sum_i M_{i,j} k_{i,N-1} \tag{2.4}$$

$$k_{i,N} = \left(\frac{1}{k_{i,0}}\right) \sum_j M_{i,j} k_{j,N-1} \tag{2.5}$$

Where N refers to the number of iterations. After that, when replacing (2.4) in (2.5), we have:

$$k_{j,N} = \sum_{j'} \widetilde{M}_{jj'} k_{j',N-2} \tag{2.6}$$

Where:

$$\widetilde{M}_{jj'} = \sum_i (M_{i,j} M_{i,j'})/(k_{j,0} k_{i,0}) \tag{2.7}$$

Equation (2.7) is satisfied when $k_{j,N} = k_{j,N-2} = 1$. This condition is met by the eigenvector of $M_{jj'}$ associated with its largest eigenvalue. However, the first eigenvector does not provide informative results as all its values are 1, making it a unit vector. Hence, the complexity measure relies on the eigenvector of $M_{jj'}$ associated with its second largest eigenvalue, which captures the majority of the variance in the original data. Consequently, the Economic Complexity Index (ECI) is defined as follows:

$$ECI = \frac{\vec{K} - \left\langle \vec{K} \right\rangle}{stdev(\vec{K})} \tag{2.8}$$

Where K is the eigenvector associated with the second largest eigenvalue of $M_{jj'}$, the operator $\langle\rangle$ denotes the mean, and *stdev* represents the standard deviation.

The Product Complexity Index (PCI) is measured using a similar approach. It involves replacing (2.5) in (2.4). The PCI is derived from the eigenvector (Q) associated with the second largest eigenvalue of the matrix $M_{ii'}$. Formally, it can be expressed as follows:

$$PCI = \frac{\vec{Q} - \left\langle \vec{Q} \right\rangle}{stdev(\vec{Q})} \tag{2.9}$$

2.3.2 Relatedness Measure

Hidalgo et al. (2007) introduced the concept of proximity, which plays a crucial role in understanding the relationship between economic activities. Proximity allows for the quantification of the distance between different economic activities, enabling the identification of products that share common or distinct capabilities in their production processes. This concept also extends to measuring the proximity or distance between the productive structures of different locations and specific goods. The notion of proximity has also been termed as *closeness* by Neffke, Henning and Boschma (2011) and *relatedness* by Boschma, Balland and Kogler (2015). These conceptual advancements have laid the foundation for numerous studies focused on characterizing the diversification patterns of countries and regions.

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 by multiple countries, it implies that they share common production capabilities and are therefore considered 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:

$$\phi_{i,f} = min\{P(RCA_i|RCA_f), P(RCA_f|RCA_i)\} \tag{2.10}$$

In this expression, for a location j:

$$RCA_{i,j} = \begin{cases} 1, & \text{if } RCA_{i,j} \geq 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.11}$$

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, Heimeriks and Balland (2014):

$$RelatednessDensity_{i,j} = \frac{\sum_{f \in j, f \neq i} \phi_{i,f}}{\sum_{f \neq i} \phi_{i,f}} \tag{2.12}$$

Equation (2.12) presents 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 1. 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 1 means that all goods related to product i are competitively produced by country j.

2.3.3 Data

To assess the diversification process of Brazilian micro-regions, 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; FRANCOSO; BOSCHMA; VONORTAS, 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 location quotient instead of the RCA and adapted the concept of proximity based on the co-location of productive activities among Brazilian micro-regions. Formally, expressions (2.1) and (2.10) are constructed as follows:

$$LQ_{i,j} = \frac{Emp_{i,j}}{\sum_i Emp_{i,j}} \div \frac{\sum_j Emp_{i,j}}{\sum_j \sum_i Emp_{i,j}} \tag{2.13}$$

$$\phi_{i,f} = min\{P(LQ_i|LQ_f), P(LQ_f|LQ_i)\} \tag{2.14}$$

In equation (2.13), the quantity exported is replaced by employment in productive activity i and micro-region j. In equation (2.14), 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 micro-regions and, consequently, facilitates the analysis of the diversification proposed in this chapter.

The main source of data for this study is the Annual Social Information Report (RAIS), which is organized by the Ministry of Labor and Employment. This database contains administrative records that are mandatory for formal employment establishments in Brazil to declare. From the RAIS, it is possible to obtain the number of formal jobs by municipality (and region) and by sector of economic activity.

Furthermore, we have grouped the economic activities according to the class (6-digits) of the National Classification of Economic Activities (CNAE) proposed by the Brazilian Institute of Geography and Statistics (IBGE). The choice of grouping based on the CNAE class is justified because it represents an intermediate level of classification, which is not overly aggregated and allows for a more comprehensive understanding of the activities in Brazil. It is also not overly specific, as the analysis conducted in this study does not require a detailed breakdown of activities within the productive structure of the regions.

The chosen territorial unit for analysis is that of micro-regions[2], as defined by IBGE (IBGE, 1990). This choice is justified by the need to capture a more accurate spatial proximity between establishments. At more aggregated levels, such as meso-regions or states, the average geographic area is larger, leading to potential misrepresentation of spatial distances between establishments. On the other hand, analyzing at a more disaggregated level, such as municipalities, may result in missing data for specific activity sectors. Therefore, analyzing at the micro-regional level allows for a more representative understanding of the activities between locations, without the drawbacks associated with more aggregated regions.

Such choices enable the construction of a database that includes 558 Brazilian micro-regions and 670 productive activities classified according to the CNAE class. The data availability for this classification covers the years from 2006 to 2021. While some graphs will refer to this period, the econometric estimates will encompass specific periods within this interval, which will be further explained below.

2.3.4 Econometric Specifications

The model specification follows the pattern tested by Neffke, Henning and Boschma (2011) and replicated in several other articles[3]. 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 definitions of these variables are provided below.

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 micro-region 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 micro-region 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 micro-region 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:

[2] Micro-regions 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] Such references were discussed in the second section of this chapter.

$$Entry_{j,i,t} = I(i \notin PF(j,t) \cap i \in PF(j,t+5)) \tag{2.15}$$

$$Exit_{j,i,t} = I(i \in PF(j,t) \cap i \notin PF(j,t+5)) \tag{2.16}$$

The choice of 5-year intervals follows the approach adopted by Boschma, Heimeriks and Balland (2014), Boschma, Balland and Kogler (2015) and Freitas (2019). Within the available period from 2006 to 2021, we selected two 5-year periods between 2009 and 2019 (2009-2014 and 2014-2019). The decision not to include the years 2020 and 2021 aims to avoid potential biases caused by the economic slowdown resulting from the coronavirus pandemic. This results 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 micro-regions 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:

$$\begin{aligned} Entry_{j,i,t} = \beta_1 RelatednessDensity_{j,i,t-5} + \beta_2 PCI_{i,t-5} + \\ \beta_3 Regions_{j,t-5} + \beta_4 Activities_{i,t-5} + \phi_j + \psi_i + \varepsilon_{j,i,t} \end{aligned} \tag{2.17}$$

$$\begin{aligned} Exit_{j,i,t} = \beta_1 RelatednessDensity_{j,i,t-5} + \beta_2 PCI_{i,t-5} + \\ \beta_3 Regions_{j,t-5} + \beta_4 Activities_{i,t-5} + \phi_j + \psi_i + \varepsilon_{j,i,t} \end{aligned} \tag{2.18}$$

Where $Regions_{j,t-5}$ is the vector of variables used to control for observable characteristics that vary over time in Brazilian micro-regions. Similarly, $Activities_{i,t-5}$ is a vector of variables that summarize the characteristics of productive activities. The fixed effects are represented by ϕ_j for regions and ψ_i for activities. Finally, $\varepsilon_{j,i,t}$ represents the residuals.

The variables of interest for the analyses conducted in this chapter are *Relatedness Density* and *PCI*. However, we also include additional variables to control for specific characteristics of regions and activities, following the formulation by Boschma, Balland and Kogler (2015). The table below summarizes the control variables used.

Table 1 – Control Variables

Variables	Description	Operation	Source	Vector
GDPpc	Gross Domestic Product per capita	$\frac{Gross\ Domestic\ Product_j}{Population_j}$	IBGE	$Regions_j$
Population	Micro-regions population	$Population_j$	IBGE	$Regions_j$
Region Productivity	Quotient between micro-region's salary mass and the number of formal jobs	$\frac{Salary\ Mass_j}{Employment_j}$	RAIS	$Regions_j$
Region HC	Percentage of workers with at least an incomplete undegraduate degree in the micro-region	$\frac{High\ Skilled\ Employment_j}{Employment_j} \times 100$	RAIS	$Regions_j$
Incentives	Percentage of municipalities with incentives to attract entrepreneurial activities	$\frac{Municipalities\ with\ Incentives}{Total\ of\ Municipalities} \times 100$	MUNIC[a]	$Regions_j$
Credit	Volume of credit operations per capita	$\frac{Volume\ of\ credit\ operations_j}{Population_j}$	ESTBAN[b]	$Regions_j$
Diversity	Diversity measured by Shannon (1948)'s index	$e^{-\sum_{i=1}^{I} \frac{Employment_{j,i}}{Employment_j} ln\left(\frac{Employment_{j,i}}{Employment_j}\right)}$	RAIS	$Regions_j$
Sector Size	Average number of employees per establishment	$\frac{Employment_i}{Establishment_i}$	RAIS	$Activities_i$
Sector HC	Percentage of workers with at least an incomplete undegraduate degree in the activity	$\frac{High\ Skilled\ Employment_i}{Employment_i} \times 100$	RAIS	$Activities_i$
Sector Productivity	Quotient between activity's salary mass and the number of formal jobs	$\frac{SalaryMass_i}{Employment_i}$	RAIS	$Activities_i$
CL	Spatial concentration of activity measured by the Coefficient of Localization (FLORENCE, 1948)	$\frac{1}{2}\sum_j \left\lvert \frac{Employment_{i,j}}{Employment_i} - \frac{Employment_j}{Employment} \right\rvert$	RAIS	$Activities_i$

[a]Municipal Basic Information Survey (MUNIC) conducted by IBGE
[b]Monthly Banking Statistics by municipality (ESTBAN) provided by the Central Bank

Source: own elaboration.

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 (CL). The inclusion of the diversity variable aligns with the literature that aims to understand the influence of a diverse local economy in attracting new industries (GLAESER et al., 1992; HENDERSON; KUNCORO; TURNER, 1995). This variable reflects the effects of Jacobs externalities, which highlight the positive impacts of diversity on regional growth. Furthermore, the Coefficient of Localization (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. These control variables help to account for the specific characteristics and dynamics of each region and activity, allowing for a more accurate assessment of the influences of Relatedness Density and PCI on sectoral entry and exit.

Finally, it is important to note that regressions 2.17 and 2.18 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 that presents the descriptive analysis of the data. By considering the complexity of each region, we aim to capture the heterogeneity in diversification dynamics and better understand the role of the complexity.

2.3.4.1 Estimation Strategy

Finally, the estimation strategy for equations (2.17) and (2.18) follows previous research. Since the dependent variables being tested can only take values of 1 or 0, representing the occurrence or non-occurrence of an event, binary response models must be employed, such as linear probability models (LPM), logit, and probit models. As defined by Wooldridge (2010), the focus of these models is on the probability of response:

$$p(x) \equiv P(y = 1|\mathbf{x}) = P(y = 1|x_1...x_K) \tag{2.19}$$

Where the partial effect of x_k on the response probability is defined as:

$$\frac{\partial P(y = 1|\mathbf{x})}{\partial x_k} = \frac{\partial p(\mathbf{x})}{\partial x_k} \tag{2.20}$$

The simplest and most commonly used model to estimate this probability is the linear probability model (LPM), which is based on the assumptions of ordinary least squares (OLS) method. However, when using LPM, estimating the coefficients that reflect the partial effect of each covariate often presents problems due to the presence of heteroscedasticity. To address this issue, heteroskedastic-robust standard errors are employed (WOOLDRIDGE, 2010). Therefore, the estimates will always be computed using robust and clustered standard errors at the micro-region and activity level. To ensure the robustness of the results, estimates will also be obtained using the logit and probit index models.

So, these models take the following form:

$$P(y = 1|\mathbf{x}) = F(\mathbf{x}\boldsymbol{\beta}) \tag{2.21}$$

Where $\mathbf{x}$ is a vector of dimension $1 \times K$ and $\boldsymbol{\beta}$ has the dimension of $K \times 1$. The function $F(\cdot)$ typically takes values within a unit interval: $0 < F(z) < 1$, $\forall z \in \Re$. Moreover, in most applications of the method, the function $F(\cdot)$ represents a cumulative distribution function and can be derived from a latent variable model, where:

$$y^* = \mathbf{x}\boldsymbol{\beta} + e, \qquad y = 1[y^* > 0] \tag{2.22}$$

The function $1[\cdot]$ is the indicator function and represents that, if $y^* > 0$, $y = 1$. Assuming that $F(\cdot)$ is the cumulative distribution function of e, and is symmetric about zero, we can state that:

$$P(y = 1|\mathbf{x}) = P(y^* > 0|\mathbf{x}) = P(e > -\mathbf{x}\boldsymbol{\beta}|\mathbf{x}) = 1 - F(-\mathbf{x}\boldsymbol{\beta}) = F(\mathbf{x}\boldsymbol{\beta}) \tag{2.23}$$

Since $1 - F(-z) = F(z)$. Therefore, the difference between the logit and probit models lies in the function assumed to represent $F(\cdot)$. The probit model is derived when e is assumed to have a *standard normal distribution*, while the logit model is derived when e is assumed to have a *standard logistic distribution*. However, the primary interest of these models is to interpret the effect of the independent variables on the response probability. To do so, we need to understand how to interpret β_k. Assuming that the covariate x_k is continuous, we have:

$$\frac{\partial P(y = 1|\mathbf{x})}{x_k} = f(\mathbf{x}\boldsymbol{\beta})\beta_k \tag{2.24}$$

So, the interpretation of the partial effect of x_k on the response probability depends on the specific point in the distribution, as $f(\mathbf{x}\boldsymbol{\beta})$ is part of expression (2.24). This issue arises from the non-linearity of these models, which hinders interpretation in the same manner as in linear models (LPM) where linearity is assumed. One approach to address this is by calculating the *average marginal effects* (AME), which synthesizes the marginal effects by taking the average value of $f(\mathbf{x}\boldsymbol{\beta})\beta_k$ across all observations in the database. In this way, we can estimate and interpret the effect of each independent variable. With the measures defined, the constructed database, and the model specification, the next section will focus on conducting a descriptive analysis of the data.

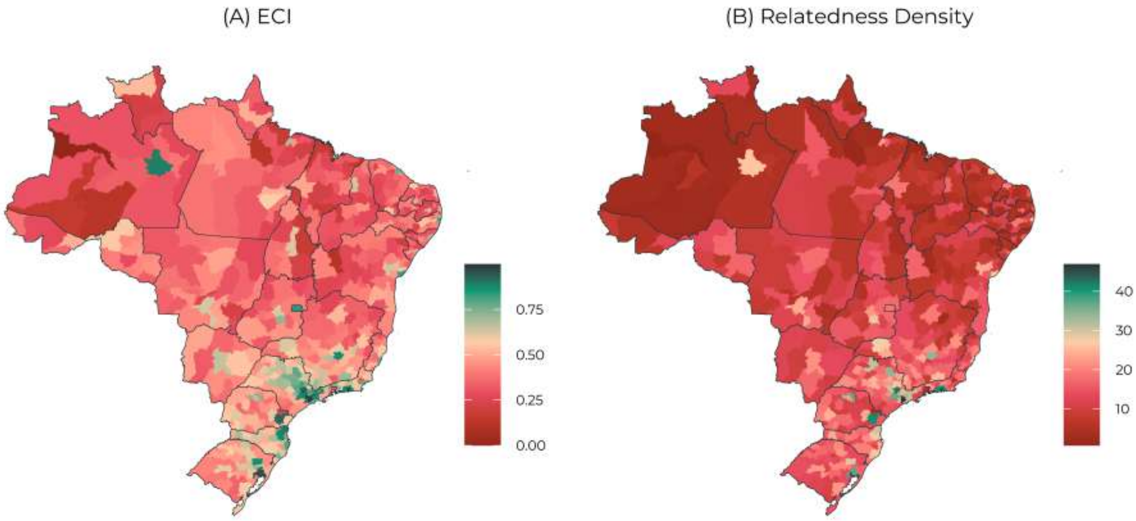
2.4 Descriptive Data Analysis

2.4.1 Economic Complexity Indicators

Table 2 presents the average complexity of activities, as measured by PCI, from 2006 to 2021 across all micro-regions in Brazil. The activities are categorized based on their PCI scores, and the table highlights the ten highest and ten lowest scoring sectors. The most complex sectors predominantly include manufacturing industry, financial activities, and other professional services. Conversely, the less complex sectors primarily consist of agricultural activities, commerce, and public administration. Public administration ranks as the least complex activity due to its widespread presence and representation across various micro-regions.

Figure 1, on the other hand, illustrates the results for measuring the ECI and Relatedness Density in Brazilian micro-regions. The averages from 2006 to 2021 are considered in both cases. The figure reveals that high complexity is predominantly concentrated in the state capitals. However, there are exceptions in the southeastern and southern regions where certain states exhibit higher complexity beyond these limits. Nevertheless, the concentration of high complexity appears to be primarily limited to large urban centers (BALLAND et al., 2020). Regions that specialize in the production of primary goods tend to display lower levels of complexity, as observed in most states in the Midwest and North regions.

Figure 1 – Average ECI and Average Relatedness Density (2006-2021)



Source: own elaboration.

The average Relatedness Density provides another perspective for analysis. A higher average density indicates that activities outside the region's existing portfolio are closer to the set of existing activities. In other words, this average represents a measure of a region's potential to diversify into new economic activities. However, as shown in Figure 1, this potential is primarily concentrated in the Southeast and South regions, particularly in the state of São Paulo. The other regions display much lower levels of density. Additionally, this figure reinforces the pervasiveness of inequality in the regional diversification process, as the potential for diversification is highly unequal between regions and reflects the greater difficulty that most regions encounter in diversifying.

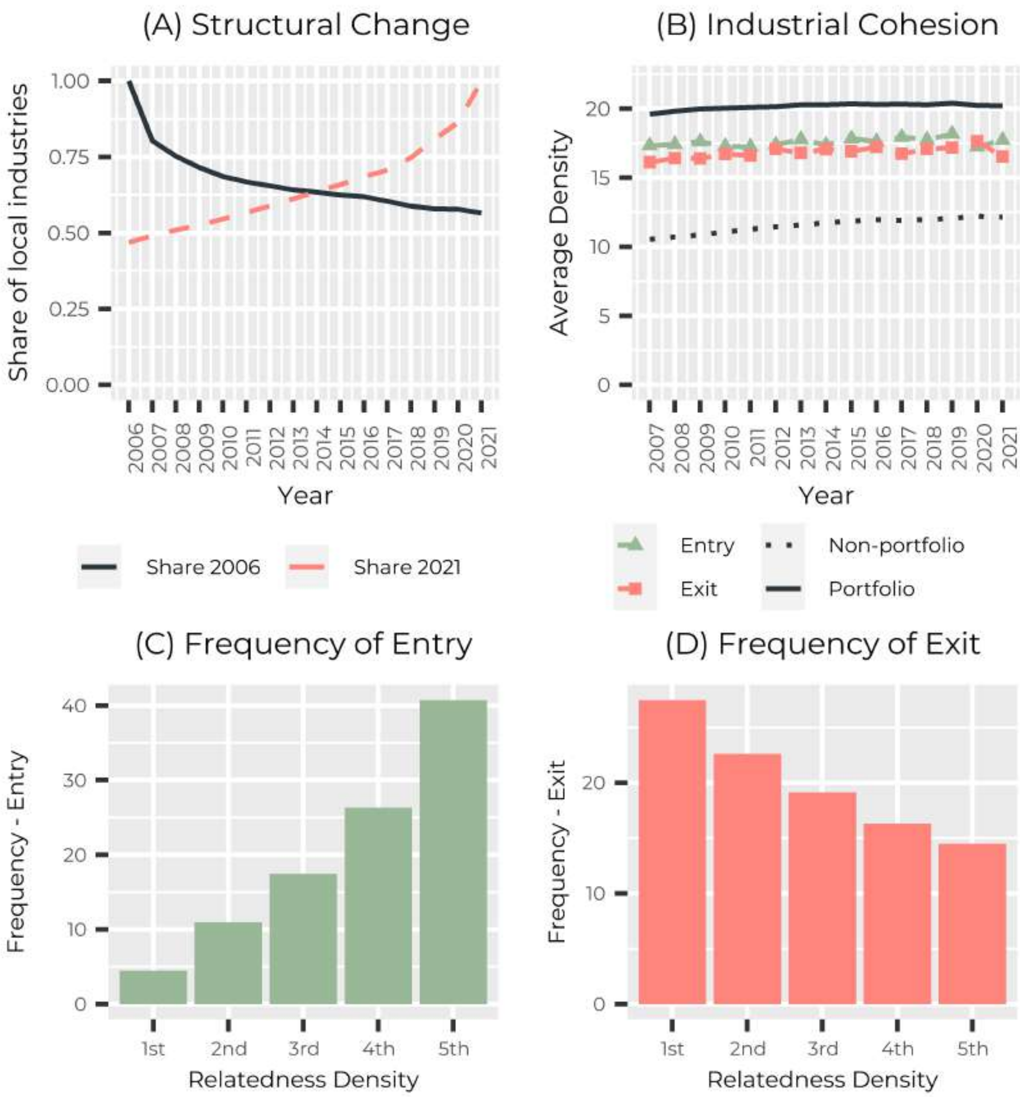
Table 2 – Average PCI (2006-2021)

CNAE Section	**CNAE Section Description**	**CNAE Class**	**CNAE Class Description**	**PCI**
C	Manufacturing industries	30504	Manufacturing of combat military vehicles	3,432
K	Financial, insurance, and related services activities	64506	Capitalization societies	3,100
M	Professional, scientific, and technical activities	70107	Company headquarters and local administrative units	2,993
K	Financial, insurance, and related services activities	64107	Central bank	2,569
C	Manufacturing industries	20941	Manufacturing of catalysts	2,484
C	Manufacturing industries	28143	Manufacturing of compressors	2,454
K	Financial, insurance, and related services activities	65120	Non-life insurance	2,428
K	Financial, insurance, and related services activities	64310	Multiple-purpose banks without commercial portfolio	2,365
K	Financial, insurance, and related services activities	64336	Development banks	2,178
C	Manufacturing industries	28241	Manufacturing of air conditioning appliances and equipment	2,177
...	...	...	...	...
A	Agriculture, livestock, forestry production, fishing, and aquaculture	01512	Cattle farming	-1,789
C	Manufacturing industries	23427	Manufacturing of non-refractory ceramic products for structural use in construction	-1,791
G	Trade; and repair of motor vehicles and motorcycles	47547	Specialized retail trade of furniture, bedding, and lighting articles	-1,835
A	Agriculture, livestock, forestry production, fishing, and aquaculture	01156	Soybean farming	-1,847
B	Extractive industries	07243	Precious metal ore mining	-1,882
G	Trade; and repair of motor vehicles and motorcycles	47121	Retail trade of goods in general, predominantly food products	-1,887
A	Agriculture, livestock, forestry production, fishing, and aquaculture	02209	Forest production - native forests	-1,947
A	Agriculture, livestock, forestry production, fishing, and aquaculture	01351	Cocoa farming	-1,968
H	Transportation, storage, and postal services	53105	Postal services	-1,991
O	Public administration, defense, and social security	84116	Public administration in general	-2,151

2.4.2 Regional Diversification in Brazil

Following the presentation of the average indicators between 2006 and 2021, it is crucial to discuss the diversification process over the years. Similar to the analyses conducted by Neffke, Henning and Boschma (2011), Essletzbichler (2015), Freitas (2019), Figure 2 presents the process of structural change by comparing the portfolio of micro-regions over the period. Additionally, it provides insights into the industrial cohesion of the productive structure of these regions based on the CNAE classes.

Figure 2 – Diversification in Brazilian micro-regions



Source: own elaboration.

Figure 2A provides a comparison of the productive structure composition of regions in 2006 and 2021, allowing for an assessment of the structural changes that occurred during this period. The continuous line represents the evolution of the share of productive activities that were present in the micro-regions' portfolios in 2006 ($LQ \geq 1$ in 2006) and their participation in subsequent years. Conversely, the dashed line represents the evolution of the participation of activities that were part of the micro-regions' productive structure in 2021 ($LQ \geq 1$ in 2021) and their share since 2006. The figure reinforces Freitas (2019)'s argument that Brazilian micro-regions have undergone significant structural changes, even at the level of CNAE classes. Only 56% of the activities that were competitively produced by the micro-regions in 2006 remained in their portfolios in 2021. Conversely, only 47% of the activities that constituted their productive structures in 2021 were part of their portfolios in 2006.

Figure 2B provides additional information on the industrial coherence of Brazilian micro-regions between 2007 and 2021. The figure consists of four lines: (i) the solid line represents the average density of activities included in the micro-regions' productive structure ($LQ \geq 1$); (ii) the dotted line shows the activities not included in the portfolio ($LQ < 1$); (iii) the dashed line with triangles shows the average density of activities that became part of the micro-regions' productive structure from one year to another; (iv) the dashed line with squares shows the average density of activities that were no longer part of the portfolio from one year to the next. Therefore, Figure 2B indicates that the productive structure of Brazilian micro-regions is cohesive, where cohesion means that the average density of industries in the portfolio is higher than that of absent ones (NEFFKE; HENNING; BOSCHMA, 2011). Furthermore, the level of coherence has remained relatively stable over the years.

The other lines provide interpretations of how diversification occurs in micro-regions in general. The entry line is consistently positioned above the dotted line (non-portfolio), indicating that new industries are more closely related to existing activities within the portfolio. However, the exit line is also above the dotted line, suggesting that the activities leaving the micro-regions' portfolio are not entirely disconnected from the existing structure. Additionally, in contrast to Freitas (2019), the entry line always overlaps with the exit line, indicating that the average density of sectors entering the portfolio is consistently higher than that of sectors leaving. Finally, it is important to note that both the entry and exit lines are distant from the solid line, indicating that the entry decreases industrial coherence while the exit increases it.

To provide further support for the ongoing analyses, it is important to evaluate the frequency of entry and exit of activities in the productive structure of the regions based on density, following the approach of Neffke, Henning and Boschma (2011). Figures 2C and 2D illustrate this frequency by density quintiles for the final years of the analysis, specifically 2014 and 2019, using the reference periods for the econometric tests (2009-2014-2019). The dynamics of the graphs become evident as the value of density increases. Among entry events, the frequency ranges from 4.5% in the first quintile to 40.7% in the last quintile. In contrast, for exit events, the frequency starts at 27.5% in the first quintile and gradually decreases to 14.5% in the last quintile.

As previously mentioned, out of the 647,801 activities in the respective micro-regions that had the potential to become part of the productive portfolios ($LQ < 1$). However, 32,223 entry events were identified. This means that the average probability of a new productive activity entering the regions' structures is 32,223/647,801 = 5%. Similarly, during the initial years of the studied period, 2009 and 2014, there were 99,919 opportunities for activities to exit the regions' portfolio ($LQ \geq 1$) and 27,616 exit events were observed. Consequently, the average probability of an activity leaving was 27,616/99,919 = 38%. These percentages closely align with the findings of Freitas (2019).

2.4.3 Complexity Groups

As mentioned earlier, 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 micro-regions:

i) *Low complexity*: micro-regions with an ECI up to 0.25.

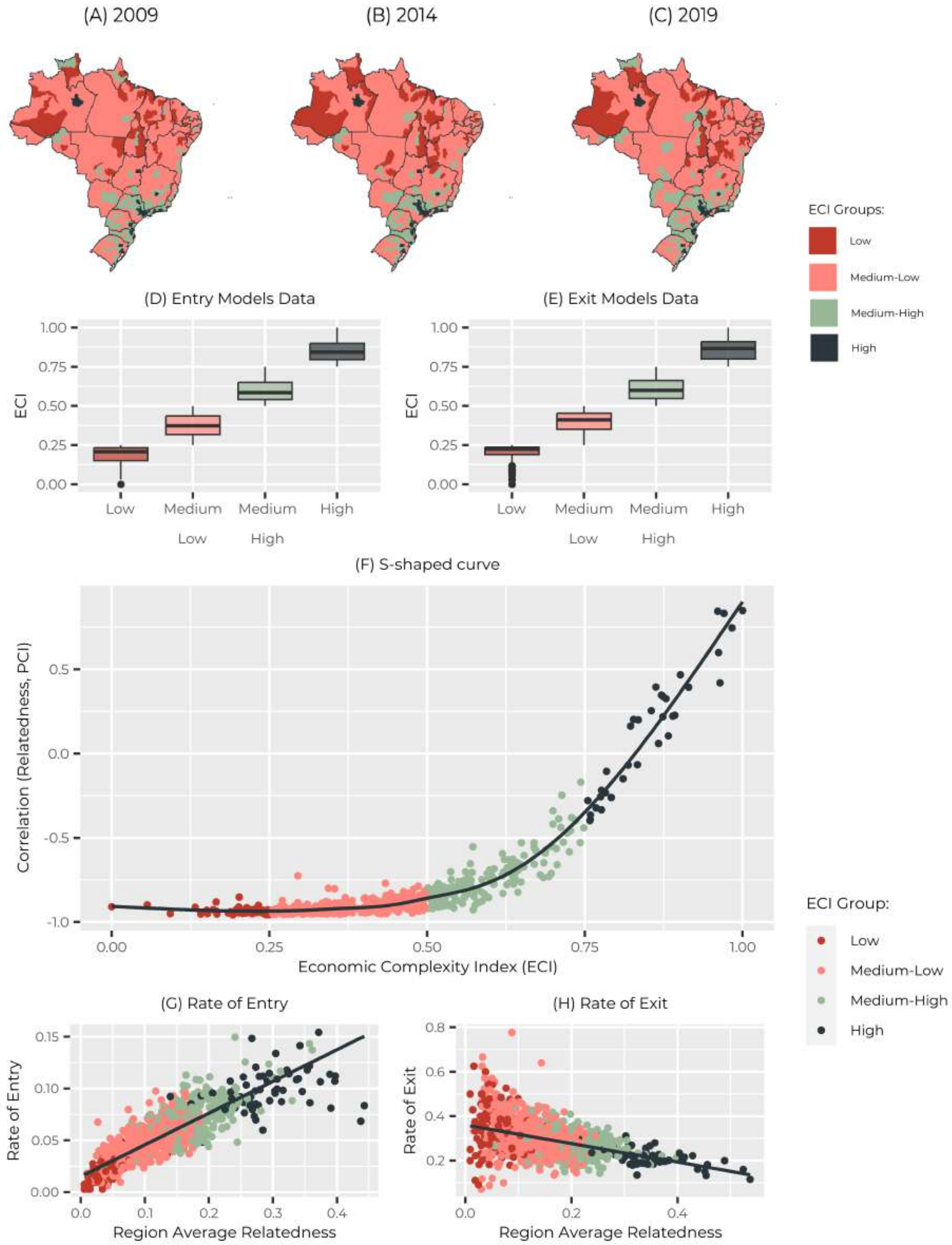
ii) *Medium-low complexity*: micro-regions with an ECI between 0.25 and 0.50.

iii) *Medium-high complexity*: micro-regions with an ECI between 0.50 and 0.75.

iv) *High complexity*: micro-regions with an ECI above 0.75.

Figures 3A, 3B and 3C show the distribution of micro-regions across these complexity groups for the reference years of analysis. The configuration of the groups appears to be consistent and stable across all years. Micro-regions with high and medium-high complexity are primarily concentrated in the South and Southeast regions, particularly around major urban centers. Micro-regions with low and medium-low complexity are located in more inland regions as well as in the North, Northeast, and Midwest regions.

Figure 3 – Complexity groups



Source: own elaboration.

Figures 3D and 3E illustrate the distribution of ECI within each complexity group, with separate graphs for entry models and exit models. The reason for this separation is that these models represent different sets of activities. In the entry models, the productive activities that were already specialized in the initial period ($LQ \geq 1$) are excluded, while in the exit models, the non-specialized activities ($LQ < 1$) are removed from the database. The distribution within each group reveals distinct patterns. Among low complexity micro-regions, ECI values are predominantly concentrated around 0.25, with some outliers at the lower end of the range, close to 0. For medium-low complexity micro-regions, the distribution tends to approach 0.50, mainly in the case of exit models. Similarly, medium-high complexity micro-regions also have a distribution closer to 0.50. Finally, highly complex micro-regions tend to have ECI values close to the lower limit of the range, around 0.75.

For the purpose of testing the hypotheses in this study, it is also crucial to visualize how the S-shaped curve would appear for the micro-regions. This curve represents the relationship between the complexity of regions and the proximity of their productive structures to new complex activities. It serves as a significant analytical tool for understanding the inherent inequality in the process of related diversification. Figure 3F presents the application of this curve to Brazilian micro-regions. Initially, it can be observed that the S-shape, commonly observed in applications for different countries (HARTMANN et al., 2020) or European regions (PINHEIRO et al., 2022), is not as prominent for the micro-regions under study. Instead, the curve appears to follow an exponential pattern. The lack of highly complex regions with structures closely aligned to new complex activities results in a different curve shape. The less complex groups (ECI between 0 and 0.50) are situated far from the more complex activities, while the groups of greater complexity (ECI between 0.50 and 1.00) gradually exhibit increased proximity to these activities.

The distinct configuration presented in Figure 3F allows for a different interpretation compared to the analysis conducted by Pinheiro et al. (2022) in the context of European regions' diversification. The authors argue that regional development in Europe can be characterized by two extreme stages, with less complex regions being close to simple activities and highly complex regions being close to more complex activities. However, in the case of Brazil, the intermediate stage of development stands out, featuring a significant number of complex micro-regions that exhibit varying levels of proximity to new complex activities. In the Brazilian context, regional development is primarily composed of two stages. Firstly, there are regions where an increase in complexity does not necessarily result in a significant reduction in the distance from the productive structures of less complex activities (ECI between 0 and 0.50). Secondly, there are other regions in an intermediate stage, where even small increments in complexity can lead to substantial increases in proximity to complex activities (ECI between 0.50 and 1.00). Thus, the Brazilian regional development exhibits a more nuanced pattern, with a notable emphasis on the intermediate stage characterized by varying levels of proximity to 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, Balland and Kogler (2015), Figures 3G and 3H show, respectively, the entry and exit rates of activities in the portfolio of micro-regions 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 3G 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 3H), there is a negative relationship with the average density of the regions, but a weaker correlation, mainly due to the less complex groups. The next section, in turn, presents the empirical tests.

2.5 Econometric Tests

Table 3 presents the basic statistics for the dependent and independent variables used in the models. Understanding the distribution of these variables is crucial for a detailed interpretation of the coefficients. It is worth noting that several covariates range between 0 and 1, reflecting participation levels, such as human capital and incentives variables.

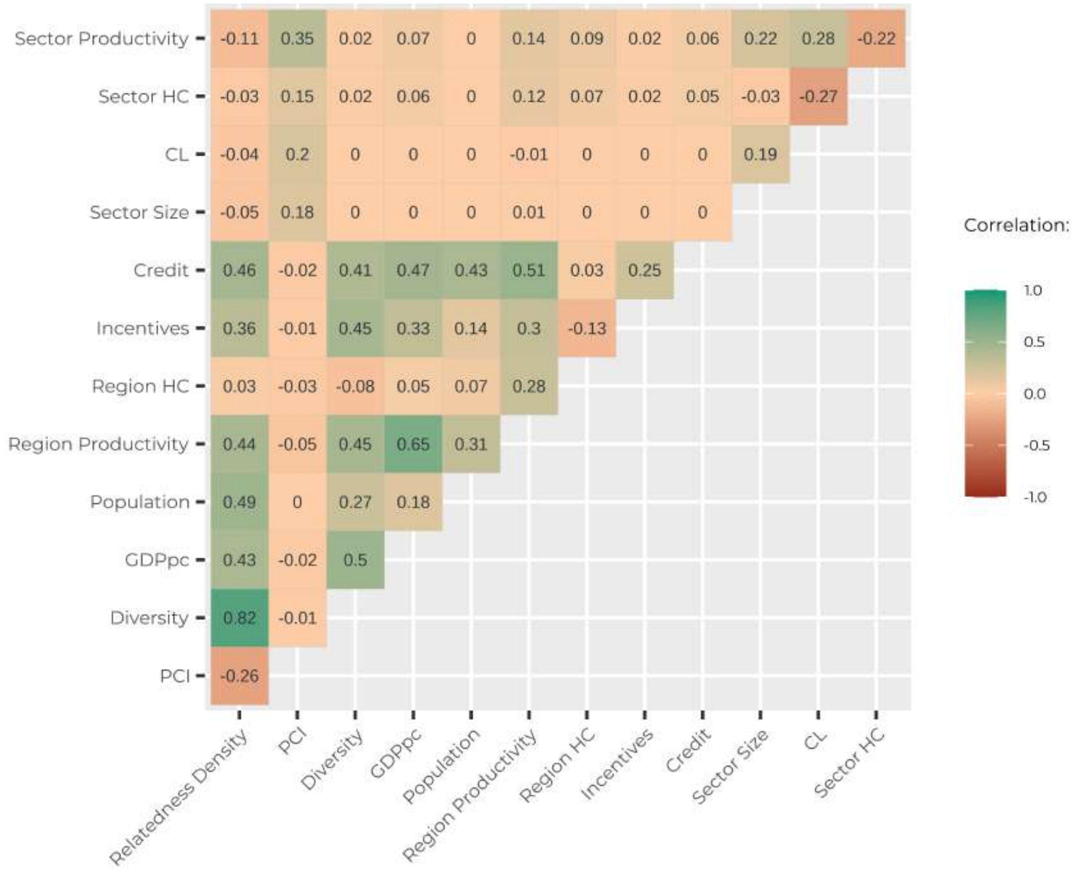
Figure 4 illustrates the strength of the correlation among these variables. Correlations are in most cases positive, indicating some degree of relationship, although they are generally weak. The strongest correlations are observed between Diversity and Relatedness (0.82) and between Regional Productivity and GDP per capita (0.65). This can be explained by the fact that more diversified economies possess a 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 3 – Basic Statistics

Variables	**Observations**	**Mean**	**Std. Deviation**	**Minimum**	**25%**	***Median***	***75%***	***Maximum***
Entry	647,801	0.05	0.22	0.00	0.00	0.00	0.00	1.00
Exit	99,919	0.28	0.45	0.00	0.00	0.00	1.00	1.00
Relatedness Density	1,121,580	0.13	0.08	0.00	0.06	0.11	0.17	1.00
PCI	1,121,580	0.38	0.17	0.00	0.25	0.36	0.49	1.00
Diversity	1,121,580	4.84	1.53	0.30	3.76	5.22	6.00	7.55
GDPpc (R$ 1.000)	1,121,580	19.93	16.00	2.69	9.33	15.81	26.04	251.68
Population	1,121,580	361,036.74	920,973.48	2,884	105,299	180,069	319,281	15,041,894
Region Productivity	1,121,580	1,586.74	623.86	487.57	1,100.62	1,541.27	1,950.49	5,455.30
Region HC	1,121,580	0.45	0.09	0.08	0.39	0.45	0.50	0.77
Incentives	1,121,580	0.59	0.26	0.00	0.40	0.62	0.79	1.00
Credit	1,121,580	4,521,381.87	7,581,362.04	0.00	1,255,196.38	2,896,716.00	6,205,963.91	195,691,980.92
Sector Size	1,121,580	50.99	164.12	1.00	8.12	16.78	37.71	3,113.39
CL	1,121,580	0.54	0.22	0.09	0.36	0.54	0.72	1.00
Sector HC	1,121,580	0.47	0.15	0.02	0.38	0.49	0.58	0.93
Sector Productivity	1,121,580	2.788,53	2,382.38	552.68	1,473.77	2,104.37	3,283.55	31.727,54

Source: own elaboration.

Figure 4 – Correlogram



Source: own elaboration.

Table 4 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 micro-regions; regression (3) considers only control variables for activities; and regressions (4), (5), and (6) are the final models specified in equation (2.17).

Table 5 presents the results of the final model considering the groups of micro-regions by complexity. Since the estimates using the three estimation strategies (OLS, logit, and probit) are similar and consistent with each other, we focus on presenting the results of the logit model in this section[4]. The corresponding estimates using OLS and probit can be found in Annex A, which ensures the robustness of the logit model. Tables 6 and 7 follow the same approach, but for the dependent variable *Exit*.

The results presented in Table 4 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.

4 Furthermore, the option for presenting the logit model for the micro-region groups is that the coefficients for PCI are slightly more significant.

Table 4 – Emergence of new activities in Brazilian micro-regions (2009-2019)

	Dependent variable:					
	Entry					
	OLS	*OLS*	*OLS*	*OLS*	*Logit*	*Probit*
	(1)	(2)	(3)	(4)	(5)	(6)
Relatedness Density	0.409***			0.646***	9.839***	4.931***
	(0.018)			(0.029)	(0.626)	(0.285)
PCI	−0.098***			−0.049**	−2.235***	−1.045***
	(0.016)			(0.019)	(0.426)	(0.205)
Diversity		0.012***		−0.005***	0.055*	0.020
		(0.001)		(0.001)	(0.031)	(0.014)
Log(GDPpc)		0.001		−0.002	−0.071	−0.024
		(0.002)		(0.002)	(0.053)	(0.024)
Log(Population)		0.008***		−0.010***	−0.218***	−0.097***
		(0.001)		(0.002)	(0.039)	(0.018)
Region Productivity		−0.011***		−0.004	−0.103	−0.055
		(0.003)		(0.003)	(0.091)	(0.041)
Region HC		0.007		−0.026***	−0.495**	−0.251**
		(0.006)		(0.008)	(0.221)	(0.099)
Incentives		0.0002		0.003	0.061	0.033
		(0.002)		(0.002)	(0.053)	(0.024)
Credit		0.000		−0.000**	−0.000*	−0.000*
		(0.000)		(0.000)	(0.000)	(0.000)
Sector Size			−0.007***	−0.004***	−0.175***	−0.076***
			(0.001)	(0.001)	(0.029)	(0.013)
CL			−0.090***	−0.084***	−2.047***	−0.982***
			(0.009)	(0.009)	(0.224)	(0.101)
Sector HC			−0.020*	0.008	0.056	0.037
			(0.012)	(0.011)	(0.245)	(0.114)
Sector Productivity			−0.011***	−0.002	−0.143*	−0.071**
			(0.002)	(0.003)	(0.076)	(0.035)
Constant	0.021**	−0.021	0.212***	0.254***	2.470***	0.893**
	(0.010)	(0.024)	(0.019)	(0.031)	(0.750)	(0.347)
Fixed effects - UF	Yes	Yes	Yes	Yes	Yes	Yes
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					0.12	0.12
Residual Std. Error	0.213	0.214	0.214	0.212		
F Statistic	260.428***	181.883***	188.327***	266.430***		

Note: *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.

The Principle of Relatedness (HIDALGO et al., 2018) is validated, indicating a relationship between Relatedness Density and the probability of entry. The proximity of micro-regions' structures to sectors not yet included in their portfolio increases the likelihood of entry, facilitating a process of related diversification. 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 (FREITAS, 2019; FRANCOSO; BOSCHMA; VONORTAS, 2022).

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, further analysis considering the complexity level of micro-regions is warranted. This process was not the primary focus of attention by Freitas (2019) and Francoso, Boschma and Vonortas (2022) and represents our main contribution.

Furthermore, several variables used to control observable characteristics of regions and activities were found to be significant in the models. Diversity, however, exhibited a change in sign from the OLS to the index models, making its effect uncertain. Nonetheless, the positive coefficients in the logit and probit models suggest that greater diversity in the local economy increases the probability of new sectors entering the regional portfolio. On the other hand, the Population variable demonstrated a significant negative effect, similar to the findings in Freitas (2019), indicating that larger micro-regions face greater challenges in attracting new sectors. GDPpc, however, was always insignificant across the models, consistent with Francoso, Boschma and Vonortas (2022). The variable representing human capital (Region HC) was found to be significant, indicating that regions with a higher proportion of educated workers tend to have a lower probability of new sectors entering their portfolio. Among the control variables for activity characteristics, sector size (Sector Size), coefficient of localization (CL), and productivity (Sector Productivity) exhibited significant and negative coefficients. This suggests that larger sectors, more spatially concentrated, and with higher productivity face greater barriers to participating in the regional structure. However, the main emphasis of this chapter is to evaluate the diversification process based on the complexity of regions, which will be demonstrated in the subsequent results.

Table 5, in turn, represents the first part of the main contribution of this article. The segmented analysis based on the complexity level of regions reveals the inherent inequality in the diversification process of Brazilian micro-regions. Across all three estimation strategies (OLS, logit, and probit), the Relatedness Density consistently shows a significant and positive effect on the probability of a new activity entering the local productive structures, irrespective of the region's complexity level. 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 micro-regions 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

Table 5 – Emergence of new activities - Logit models

	Dependent variable:				
	Entry				
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	9.839***	23.779***	11.794***	13.531***	16.059***
	(0.626)	(3.271)	(1.087)	(1.256)	(1.427)
	[0.436]	[0.437]	[0.460]	[0.838]	[1.279]
PCI	−2.235***	−6.388***	−3.584***	−0.180	0.773**
	(0.426)	(0.839)	(0.511)	(0.435)	(0.300)
	[-0.099]	[-0.117]	[-0.140]	[-0.011]	[0.062]
Diversity	0.055*	−0.079	−0.079*	−0.542***	−0.581**
	(0.031)	(0.110)	(0.044)	(0.090)	(0.252)
Log(GDPpc)	−0.071	−0.161	0.006	0.055	−0.445*
	(0.053)	(0.210)	(0.046)	(0.084)	(0.246)
Log(Population)	−0.218***	0.128**	−0.030	−0.143***	−0.799***
	(0.039)	(0.057)	(0.029)	(0.051)	(0.148)
Region Productivity	−0.103	−0.724***	−0.235**	0.062	0.389
	(0.091)	(0.274)	(0.093)	(0.150)	(0.352)
Region HC	−0.495**	−0.433	−0.499**	−0.966**	1.088
	(0.221)	(0.306)	(0.205)	(0.415)	(1.102)
Incentives	0.061	−0.204	0.087*	0.109	0.119
	(0.053)	(0.142)	(0.051)	(0.113)	(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***	−0.187***	−0.196***	−0.161***	−0.161***
	(0.029)	(0.066)	(0.035)	(0.033)	(0.043)
CL	−2.047***	−1.671***	−2.091***	−2.366***	−2.576***
	(0.224)	(0.421)	(0.254)	(0.221)	(0.257)
Sector HC	0.056	0.135	−0.345	0.554**	0.691**
	(0.245)	(0.497)	(0.306)	(0.251)	(0.348)
Sector Productivity	−0.143*	−0.115	−0.135	−0.138*	0.002
	(0.076)	(0.168)	(0.089)	(0.076)	(0.120)
Constant	2.470***	2.806	1.878***	2.296**	7.086***
	(0.750)	(1.748)	(0.722)	(1.056)	(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 R^2	0.12	0.2	0.14	0.09	0.11
Log Likelihood	−112,612.200	−5,628.153	−60,592.810	−35,442.190	−8,413.916
Akaike Inf. Crit.	225,476.300	11,492.310	121,435.600	71,130.390	17,049.830

Note: *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.
The values in square brackets [ ] represent the *average marginal effects.*

in Table 5 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 6 and 7 present the same estimates for assessing the probability of activities exiting. Comparing models (4), (5) and (6) in Table 6, as with the similar ones in the entry models reveals a consistent pattern: 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, indicating a closer connection between the regional portfolio and the activity, 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 discussed when interpreting the results of the entry models, and generally, the same variables remain significant.

Table 6 – Exit of activities in Brazilian micro-regions (2009-2019)

	Dependent variable:					
	Exit					
	OLS	*OLS*	*OLS*	*OLS*	*Logit*	*Probit*
	(1)	(2)	(3)	(4)	(5)	(6)
Relatedness Density	−0.820***			−1.334***	−7.745***	−4.507***
	(0.036)			(0.082)	(0.501)	(0.291)
PCI	0.389***			0.368***	1.848***	1.102***
	(0.045)			(0.043)	(0.238)	(0.140)
Diversity		−0.015***		0.027***	0.178***	0.102***
		(0.005)		(0.006)	(0.030)	(0.018)
Log(GDPpc)		0.001		0.005	0.015	0.012
		(0.008)		(0.008)	(0.043)	(0.025)
Log(Population)		−0.034***		0.023***	0.132***	0.081***
		(0.004)		(0.005)	(0.028)	(0.017)
Region Productivity		0.016		−0.021	−0.115	−0.066
		(0.012)		(0.016)	(0.084)	(0.051)
Region HC		0.042		0.103***	0.579***	0.338***
		(0.033)		(0.040)	(0.209)	(0.125)
Incentives		−0.001		−0.009	−0.051	−0.031
		(0.009)		(0.011)	(0.056)	(0.034)
Credit		−0.000**		0.000**	0.000***	0.000***
		(0.000)		(0.000)	(0.000)	(0.000)
Sector Size			−0.049***	−0.054***	−0.298***	−0.174***
			(0.006)	(0.006)	(0.033)	(0.019)
CL			0.214***	0.290***	1.559***	0.932***
			(0.044)	(0.039)	(0.206)	(0.123)
Sector HC			0.153***	0.112**	0.595**	0.361**
			(0.056)	(0.053)	(0.278)	(0.166)
Sector Productivity			0.028**	0.023*	0.132*	0.074*
			(0.011)	(0.013)	(0.069)	(0.041)
Constant	0.330***	0.594***	−0.012	−0.185**	−3.644***	−2.211***
	(0.029)	(0.074)	(0.081)	(0.093)	(0.501)	(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***		

Note: *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.

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 Francoso, Boschma and Vonortas (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, it is the effect of PCI that garners the most attention. The complexity of sectors is pivotal in increasing the likelihood of exiting the local portfolio, particularly in less complex groups (Low and Medium-Low). However, as the complexity of micro-regions increases, the effect of PCI diminishes and becomes insignificant. This phenomenon highlights that less complex micro-regions lack the necessary skills, knowledge, and capabilities to sustain complex activities within their structure, while more complex regions do not face the same challenge.

Table 7 – Exit of activities - Logit models

	Dependent variable:				
			Exit		
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	−7.745***	−11.517**	−12.222***	−9.621***	−10.978***
	(0.501)	(4.984)	(1.019)	(1.261)	(1.289)
	[-1.419]	[-2.095]	[-2.239]	[-1.760]	[-1.730]
PCI	1.848***	7.601***	2.687***	0.569**	−0.396
	(0.238)	(1.407)	(0.370)	(0.271)	(0.309)
	[0.338]	[1.383]	[0.492]	[0.104]	[-0.062]
Diversity	0.178***	0.196	0.350***	0.648***	0.422**
	(0.030)	(0.180)	(0.047)	(0.093)	(0.170)
Log(GDPpc)	0.015	−0.592***	−0.011	−0.001	0.277**
	(0.043)	(0.203)	(0.052)	(0.079)	(0.120)
Log(Population)	0.132***	−0.059	0.024	0.067	0.524***
	(0.028)	(0.108)	(0.032)	(0.050)	(0.114)
Region Productivity	−0.115	1.171***	−0.073	−0.268*	−0.853***
	(0.084)	(0.369)	(0.104)	(0.159)	(0.218)
Region HC	0.579***	−0.390	0.746***	0.961**	1.621*
	(0.209)	(0.649)	(0.248)	(0.454)	(0.902)
Incentives	−0.051	−0.030	−0.130*	0.053	−0.292
	(0.056)	(0.235)	(0.072)	(0.126)	(0.344)
Credit	0.000***	−0.000	0.000	0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sector Size	−0.298***	−0.305**	−0.302***	−0.285***	−0.264***
	(0.033)	(0.124)	(0.041)	(0.036)	(0.043)
CL	1.559***	1.456**	1.431***	1.957***	3.357***
	(0.206)	(0.727)	(0.247)	(0.222)	(0.290)
Sector HC	0.595**	−0.229	0.622*	0.407	0.838**
	(0.278)	(0.963)	(0.369)	(0.282)	(0.330)
Sector Productivity	0.132*	0.388	0.149	0.096	0.020
	(0.069)	(0.261)	(0.092)	(0.079)	(0.102)
Constant	−3.644***	−10.714***	−2.934***	−3.909***	−5.220***
	(0.501)	(2.753)	(0.665)	(0.984)	(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 R^2	0.07	0.15	0.10	0.06	0.07
Log Likelihood	−54,737.750	−1,658.514	−25,890.140	−20,160.710	−6,223.971
Akaike Inf. Crit.	109,727.500	3,535.029	52,030.280	40,567.430	12,669.940

Note: *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.
The values in square brackets [ ] represent the *average marginal effects.*

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 7, 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 micro-regions. 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. Table 8 summarizes these key results.

Table 8 – Summary of Entry and Exit Model Results - Logit models[1]

	Dependent variable: Entry				
	General Model	Low	Medium-Low	Medium-High	High
Relatedness Density	0.436***	0.437***	0.460***	0.838***	1.279***
PCI	−0.099***	−0.117***	−0.140***	−0.011	0.062**
	Dependent variable: Exit				
	General Model	Low	Medium-Low	Medium-High	High
Relatedness Density	−1.419***	−2.095**	−2.239***	−1.760***	−1.730***
PCI	0.338***	1.383***	0.492***	0.104**	−0.062

[1]The values presented are the *average marginal effects.*
Signif.: *p<0.1; **p<0.05; ***p<0.01

The econometric tests conducted in this chapter 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. 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. 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 1 and 2, 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.

Furthermore, the econometric results reinforce the interpretation discussed in Figure 3F, 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 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.

Even though the tests were performed based on three different specification strategies, some robustness tests were performed to ensure the consistency of the results. In the annex A are the estimates used to reinforce the analysis. First, the same tests were done for different time periods. Two 5-year windows between 2007 and 2017 were considered. Second, the way of classifying the complexity groups was changed. Instead of segmenting by ECI values, it became segmented by quartiles. Third, an alternative measure for calculating density was also considered. No longer by co-location, the reference indicator was calculated by the concept of co-occupation. Finally, in all cases the results were consistent and therefore robust.

2.6 Concluding Remarks

This chapter 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 558 Brazilian micro-regions between 2006 and 2021 yielded significant findings. Firstly, complexity in Brazil is regionally concentrated, with the highest rates observed in regions consisting of state capitals. The potential for diversification, as measured by the average regional density, is even more concentrated. The proximity to new sectors, in general, is primarily concentrated in the Southeast and South regions, particularly in the state of São Paulo. Furthermore, the assessment of industrial cohesion in Brazil indicates that the structures remained cohesive throughout the analyzed period, as the average density of portfolios consistently maintained a distinct distance from the density of activities that were not part of the local structures.

The descriptive analysis revealed that the S-shaped curve actually takes on an exponential form in the case of Brazilian micro-regions. Instead of development occurring in two distinct stages, as argued by Pinheiro et al. (2022) for European regions, 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. This characterization of regional development in Brazil underscores the challenges that low-complexity regions may encounter in diversifying their economies. 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 productive 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 proved to be crucial and played a key role in the findings presented in this chapter. 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 chapter 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. Freitas (2019) mentioned the tests only for the most complex regions, and Francoso, Boschma and Vonortas (2022) did not assess the probability of sector exit, in addition to employing a more aggregated regional classification. Therefore, the main contribution of this chapter 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. Additionally, The chapter offers a novel analysis of the S-curve in a developing country, emphasizing economic development disparities in light of complexity compared to developed 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? 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.

3 How good or how bad is the news, in terms of employment?

3.1 Introduction

The complexity of the productive structure is a relevant predictor of future economic growth (HAUSMANN et al., 2014) and employment (ROMERO et al., 2022; QUEIROZ; ROMERO; FREITAS, 2023). Therefore, the shift towards more complex sectors is considered crucial in the literature on economic complexity (HIDALGO et al., 2007). The accumulation of diverse and distinct capabilities provides economies with the opportunity to diversify and gain competitiveness across a wide range of goods, which ultimately impacts overall economic performance. However, the process of economic diversification is strongly path-dependent, and regions face limitations in their ability to diversify. This constraint is reflected in economic polarization resulting from related diversification, exacerbating regional inequalities (PINHEIRO et al., 2022; HARTMANN; PINHEIRO, 2022). This condition creates winners and losers and, among other effects, significantly impacts the job creation process in these economies.

The literature on complexity and regional inequality is still very limited. Two notable exceptions are Hartmann and Pinheiro (2022) and Pinheiro et al. (2022). The former assesses the variation in the relationship between complexity and inequality at the regional level compared to the national level. The latter examines the differences in diversification patterns among European regions. In both cases, the authors draw attention to the uneven development resulting from the process of related diversification within regions. Nonetheless, the literature still requires studies that assess the implications of these dynamics.

In this context, this chapter aims to examine the heterogeneity of local employment multipliers due to differences in regional complexity. This analysis is made possible by adapting the conceptual framework found in the literature on local multipliers. This theoretical field was transformed by the seminal contribution of Moretti (2010). In his study, Moretti examines local employment multipliers in tradable and non-tradable sectors in US cities. The innovation of this influential article lies in the proposition of a simplified conceptual and econometric framework for calculating multipliers, utilizing shift-share instruments. Subsequently, numerous analyses have followed this framework to calculate employment multipliers in regions across various countries in Europe, Japan (KAZEKAMI, 2017), China (WANG; CHANDA, 2018) and Brazil (MACEDO; MONASTERIO, 2016; LOYO; MOISÉS; MENDES, 2018; ROCHA; ARAÚJO, 2021). Therefore, this paper adapts this framework to calculate multipliers for complex and non-complex sectors. In other words, our interest stems from the need to understand *how good* or *how bad* it is to have the regional economy specialized in more or less complex activities, in terms of local employment growth.

Based on the complexity indicators developed by Hidalgo and Hausmann (2009), we divided the region's economies into two sectors: complex and non-complex. Using formal labor market data from the micro-regions of Brazil at three different time points (2009, 2014 and 2019), we examined all possible relationships between these sectors, addressing potential endogeneity issues by employing instrumental shift-share variables. In addition to using the conventional shift-share instrument proposed by Moretti and Thulin (2013), we also introduced an instrument that takes into account regional structural changes. Finally, due to the limited explanations available in the literature, we conducted a bootstrap analysis to assess the changes in employment within the same sector.

In general, the multiplier for complex industries is expected to be greater than for non-complex industries. This hypothesis is based on the reasoning that the presence of complex and less ubiquitous activities (manufacturing and modern services) in the local economy tends to create a greater demand for less complex and more ubiquitous activities (trade of food and beverages, services of cleaning, construction materials). Moreover, the analysis takes into account the differentiation of micro-regions according to their complexity levels (Low, Medium-Low, Medium-High and High). It is assumed that complex sector multipliers will be greater in regions that are already complex. These regions tend to have more established institutions, higher labor competition between sectors, and greater labor mobility. These characteristics make the labor supply more responsive to changes in the complex sector. In summary, these hypotheses are supported by the results of econometric tests conducted in the study.

The chapter is organized as follows. Section 2 briefly outlines the limitations of the existing literature on complexity and regional inequality and focuses on reviewing the literature on local employment multipliers. It is the gaps in the former and the trends in the latter that lead to the main focus of this chapter. Section 3 brings the data used and the method adapted from Moretti and Thulin (2013). Section 4 performs a descriptive analysis of the data, briefly discussing regional inequality in Brazil and the complexity classifications used. Section 5, in turn, brings the results of the econometric estimates and Section 6 concludes the chapter with final considerations.

3.2 Complexity, Regional Inequality, and Local Employment Multipliers

The literature on the Principle of Relatedness and its interaction with the economic complexity approach has provided new insights into regional inequalities. As discussed in Chapter 2, regional economic diversification tends to occur within sectors already related to existing structures, leading to uneven and divergent development between regions. In the complexity approach, this dynamic results in the emergence of winning and losing regions, where diversification into new, more complex sectors is primarily limited to regions that already possess complex productive capacities (winners). On the other hand, other regions diversify into sectors that are closely aligned with local structures but not necessarily more complex (losers). This presents a problem as the literature suggests that complexity plays a crucial role in driving income and employment growth (HAUSMANN et al., 2014; ROMERO et al., 2022).

Recent notable contributions have shed light on this inherent unequal aspect of regional economic development. Hartmann and Pinheiro (2022) and Pinheiro et al. (2022) emphasize the presence of a "feedback loop of spatial inequality" in their analysis of European regions and a "winner-take-most" dynamic when assessing regional development. These expressions are used to describe how regional diversification, in relation to specific sectors, exacerbates the disparities between poor and rich regions, specialized and diversified regions, and less and more complex regions. However, the implications of this process are still relatively understudied.

As discussed in Chapter 1, the literature examining the relationship between complexity and regional inequality is limited in some key aspects. Firstly, the majority of studies primarily focus on income inequality, often utilizing indices that assess its distribution. Few studies consider other forms of inequality, such as wage inequality (SBARDELLA; PUGLIESE; PIETRONERO, 2017). Secondly, the existing contributions have yet to fully explore the impact of these complexity differentials on other dimensions, such as their influence on local economic drivers like job creation. In summary, the current contributions are primarily centered around regressing common indices for evaluating income distribution (e.g., Gini, Theil) against complexity indicators (e.g., ECI and Fitness), thus limiting the scope of analysis. Nevertheless, in this chapter, the objective is to examine one of these implications: the magnitude of regional employment multipliers based on local complexity. The next subsection reviews the literature on the methodology that will be used to measure the multipliers.

3.2.1 Literature on Local Employment Multipliers

The methodology employed to measure regional employment multipliers is derived from the seminal article by Moretti (2010). Moretti proposed a simple conceptual framework for estimating local employment multipliers in the manufacturing sector. According to this framework, each city is viewed as a competitive economy comprising two types of goods: tradables (manufactured goods) and non-tradables (services and other industries). Tradables have a price defined at the national level, while non-tradables have a price determined locally. The framework assumes labor mobility across sectors within cities, ensuring that the marginal wage and marginal product are equal. Additionally, it considers that the supply curves for labor and housing are positively sloped. The slope of the labor supply curve depends on residents' preference for leisure and the level of labor mobility between cities, while the slope of the housing supply curve is influenced by local land use regulations. With this conceptualization, the author calculates a local multiplier for the non-tradable sector and another for the tradable sector, resulting from an exogenous increase in employment within the tradable (manufacturing) sectors.

In short, Moretti (2010) studies the implications of a permanent increase in jobs in the tradable goods sector, which can result from the arrival of a new firm or a substantial increase in demand from existing firms. According to the assumptions outlined earlier, such a shock affects the general equilibrium of prices. As a result, workers' wages and housing costs increase, unless the supply curves are perfectly elastic. This process leads to an expansion of the city's budget constraint, with more jobs and higher wages, consequently increasing demand in non-tradable sectors such as personal services, restaurants, cleaning services, and more. This represents the multiplier effect for the non-tradable sector (3.1). On the other hand, the shock also impacts the tradable sector through three distinct effects. First, the rise in local labor costs makes the existing tradable sector less competitive, as prices are not subject to the same local dynamics. Second, there may be increased demand in intermediate tradable sectors, and the extent of this impact depends on the geographic concentration of these industries. Third, agglomeration effects occur as a consequence of the initial employment shock. The combined influence of these factors gives rise to the multiplier effect for the tradable sector (3.2). Formally, Moretti (2010) proposes the following solution:

$$\Delta N_{ct}^{NT} = \alpha + \beta \Delta N_{ct}^{T} + \gamma d_t + \epsilon_{ct} \tag{3.1}$$

$$\Delta N_{ct}^{T_1} = \alpha' + \beta' \Delta N_{ct}^{T_2} + \gamma' d_t + \epsilon'_{ct} \tag{3.2}$$

Where ΔN_{ct}^{NT} and ΔN_{ct}^{T} are the change over time in the log number of jobs, respectively, in the non-tradable and tradable sectors. Moreover, $\Delta N_{ct}^{T_1}$ is the change in the log number of jobs in a randomly selected part of tradable sector and $\Delta N_{ct}^{T_2}$ represents the change in the log number of jobs in the remaining part of the tradable sector. Finaly, d_t is a dummy variable used to indicate the last period under consideration.

However, the estimation of equations (3.1) and (3.2) through OLS can lead to biased estimators due to endogeneity problems and omitted variables. Factors such as increased employment in non-tradable sectors that generate more jobs in tradable sectors, as well as unobservable time-varying shocks to local labor supply, can confound the causal effect of the shock. To address this, Moretti (2010) adopts an instrumental variable approach using a shift-share instrument (BARTIK, 1991). The instrumental variable is constructed as the average nationwide employment growth in manufacturing industries, weighted by the share of these industries in cities during the initial period. By assuming that national changes in employment are exogenous to region-specific dynamics, regression with an instrumental variable can provide unbiased estimators.

Moretti (2010) presents an alternative methodology for measuring multipliers, which diverges from the conventional approach of using input-output matrices. The author argues that estimates obtained through Input-Output analysis overlook the employment effects in non-tradable sectors and fail to capture job losses in tradable sectors resulting from increased labor costs, as well as the gains from agglomeration economies. As a solution, Moretti proposes a calculation method based on a simple regression framework. This approach offers a user-friendly implementation for policymakers to evaluate the impact of their actions. Furthermore, it has paved the way for a series of subsequent articles that utilize this tool to calculate regional multipliers, providing valuable insights for regional economic analysis and policy-making.

In Moretti (2010)'s pioneering study, he provided the initial estimates of multipliers that served as a benchmark for subsequent research. He found that for each additional job created in the tradable sector, 1.6 jobs are generated in the non-tradable sector within the same city. Moreover, Moretti (2010) argues that skilled jobs have a greater multiplier effect due to their concentration of higher wages. Specifically, he found that each additional skilled job generates 2.5 non-tradable jobs. Furthermore, the author suggests that the multiplier for the tradable sector should be relatively smaller, or potentially negative, due to the increase in labor costs associated with it. Formally, Moretti's analysis reveals that an additional job in a specific part of the tradable sector generates 0.26 jobs in the remaining part. This framework and analysis have been replicated and adapted in studies conducted for various countries, expanding the understanding of regional multipliers.

Moretti and Thulin (2013) conducted a similar exercise to assess local employment multipliers in Sweden. They made an adaptation to the instrumental variable used by excluding the reference region in the instrument's measurement to address potential violations of the exogeneity assumption due to the region under analysis being included in the calculation. The findings of Moretti and Thulin (2013) revealed a statistically significant multiplier, although smaller than that observed in the United States. Specifically, the addition of one job in the tradable sector generated between 0.4 and 0.8 jobs in the non-tradable sector. The multiplier was notably higher for skilled jobs and the high-tech industry. In terms of the tradable sector, the multiplier was closer to the range of 0.3 to 0.4, similar to the findings in the United States.

Dijk (2015) furthered the adaptations by attempting to replicate Moretti (2010)'s analysis while also calculating the multipliers using the alternative instrumental variable that excludes the reference city from the calculation. The author argues that estimates obtained with the new instrument are more robust as they align with the plausibility of exogeneity. The results of this exercise yield statistically significant but smaller multipliers. Specifically, the creation of one job in the tradable sector leads to the generation of 0.84 non-tradable jobs in the same city. In the case of skilled jobs, each additional job generates 1.46 jobs in local non-tradable sectors. These multipliers are lower than the 1.6 and 2.5 estimates previously found by Moretti (2010).

Apart from the studies conducted in the United States and Sweden, several other countries have been analyzed using a similar methodology. The literature includes contributions that examine the cases of Italy (BLASIO; MENON, 2011), Spain (GEROLIMETTO; MAGRINI, 2014), United Kingdom (FAGGIO; OVERMAN, 2014), Japan (KAZEKAMI, 2017), China (WANG; CHANDA, 2018), Mexico (HERNANDEZ; ROJAS, 2020) and Brazil (MACEDO; MONASTERIO, 2016; LOYO; MOISÉS; MENDES, 2018; ROCHA; ARAÚJO, 2021).

While not all studies specifically focus on non-tradable and tradable multipliers, they adopt the same measurement methodology to assess the employment multiplier effects in regional economies. For instance, the methodology was employed to calculate employment multipliers in the public sector (FAGGIO; OVERMAN, 2014), the creative industries (GOOS; KONINGS; VANDEWEYER, 2018), and cultural industries (GUTIERREZ-POSADA et al., 2023). Moreover, the same methodology was utilized to gauge the impact of job creation on other variables, including the unemployment rate and the total number of unemployed individuals (ROCHA; ARAÚJO, 2021).

Blasio and Menon (2011) replicated Moretti (2010)'s methodology in their study of Italy. However, their findings diverged from previous works as they did not find evidence of a significant multiplier effect in the local labor market. Specifically, the increase in employment in the tradable sector did not generate notable job creation effects in either the local non-tradable sector or other local tradable sectors. This lack of impact was consistent across different regions of Italy. The authors proposed three potential reasons to explain this phenomenon. First, they suggested that excessive regulation in the non-tradable sector could hinder competition, limiting job creation. Second, the limited wage variability in the labor market could impede a responsive labor market adjustment to changes. Lastly, they highlighted additional difficulties in labor mobility as a possible contributing factor.

Gerolimetto and Magrini (2014) encountered a similar issue when examining regions in Spain. Their study utilized data from 103 labor market areas in Spain and conducted three different estimations: OLS, instrumental variable with spatial filtering, and instrumental variable alone. Interestingly, only the OLS estimation reported a significant coefficient for the change in employment in the tradable sector, while the other two strategies yielded non-significant coefficients. Due to endogeneity concerns associated with OLS and the lack of robustness in its estimates, Gerolimetto and Magrini (2014) concluded that there was no evidence of a multiplier effect in the Spanish labor market, mirroring the findings in Italy. The authors suggested that Spain, like Italy, had less flexible labor markets compared to the United States. The lower elasticity of labor and housing supply in these countries may lead to a rapid adjustment in the general price equilibrium, reducing the multiplier effects observed.

Faggio and Overman (2014) adapted the methodology to examine the impact of public sector employment on local labor markets in the UK. Their study focused on the period between 2003 and 2007. The authors found that additional employment in the public sector led to the generation of 0.5 jobs in the construction and services sectors, while simultaneously reducing 0.4 jobs in the manufacturing sector. However, they did not observe a significant increase in employment within the overall private sector. Notably, employment growth in the public sector did contribute to a 1-to-1 increase in total employment within a given location. This adaptation by Faggio and Overman (2014) served as a reference for subsequent studies investigating similar dynamics.

The analysis for Japan's multipliers followed a slightly different approach. Kazekami (2017)'s study focused on agglomeration economies and their impact on job creation. According to the author, the attraction of new jobs in the tradable sector leads to the generation of additional jobs in the non-tradable sector, particularly when labor mobility is high. In regions with high labor mobility, the multiplier effect is prominent. However, when labor mobility is low, this multiplier effect diminishes or disappears. Kazekami (2017) argued that regions experiencing a high influx of labor are likely to have higher multipliers, highlighting the positive association between agglomeration economies and the magnitude of multipliers.

Employment multipliers in the Chinese labor market were measured by Wang and Chanda (2018). The authors adapt the strategy proposed by Moretti (2010) by incorporating a set of city characteristics into the regression model that potentially influence job growth in the non-tradable sectors. The estimation results revealed that an additional job in the manufacturing sector generates an average of 0.34 jobs in the local non-tradable sector. Importantly, this multiplier effect remained statistically significant even after accounting for factors such as neighboring area development, world market access, and geographic characteristics. Additionally, Wang and Chanda (2018) observed that the multiplier varied geographically, with higher values found in interior regions of China. This heterogeneity suggests that the impact of manufacturing employment on non-tradable job creation differs across different regions of the country.

Hernandez and Rojas (2020) conducted an analysis of the Mexican labor market, focusing on the separation between the formal and informal sectors. They adapted the econometric specification proposed by Moretti and Thulin (2013) to calculate the multipliers for the tradable sector. To account for convergence effects, the authors included the level of employment in the non-tradable sector in the initial period as an independent variable. The findings revealed that one additional job in the tradable sector leads to the creation of 1.8 to 2.6 additional jobs in the non-tradable sector, with 1 to 1.5 of those jobs being in the formal sector. Hernandez and Rojas (2020) argue that expanding employment in the manufacturing sector can also serve as a policy tool to promote formalization within the labor market.

In Brazil, three studies have examined employment multipliers. Macedo and Monasterio (2016) conducted an analysis similar to Moretti (2010) using data from 21 different economic activities across 123 meso-regions in Brazil. The authors found that for each additional industrial job, 3.78 new jobs were created in the service sector, excluding the metropolitan region of São Paulo. However, when including São Paulo, the multiplier increased to 6.58. Additionally, Macedo and Monasterio (2016) found a multiplier of 6.94 for the influence of high technology industries on local services, which aligns with previous research findings. It's important to note that the authors caution against overgeneralizing these results, as the multipliers are average estimates and may not capture the unique development experiences of each region. Factors such as sector, technology, strategy, and other local characteristics can significantly impact the effect of employment shocks.

Loyo, Moisés and Mendes (2018) conducted a study focusing on employment multipliers in the Brazilian public sector, similar to the work of Faggio and Overman (2014). The study analyzed the period of the first two terms of President Lula, from 2003 to 2010. The findings suggest a change in the multiplier effect between the two terms. During the period of a contractionary fiscal policy (2003-2006), an increase in public sector employment led to a displacement of private sector employment (negative multiplier), with approximately 0.46 private jobs being displaced for every additional public job. In contrast, during the period of an expansionary fiscal policy (2007-2010), the increase in public sector employment was complementary to private sector employment (positive multiplier), resulting in the creation of approximately 0.79 new private jobs for each additional public job.

Rocha and Araújo (2021) conducted a recent study in Brazil, building upon the previous research on job multipliers. They applied a similar econometric strategy to estimate the effects of increased industrial employment on various labor market outcomes. The findings of their study indicate that an additional job in the industrial sector, on average, leads to a reduction of 2.6 unemployed individuals and an increase of 8.4 new jobs in the non-tradable sector. This suggests that industrial employment growth has a positive impact on reducing unemployment and generating employment opportunities in other sectors of the economy. The study also found supporting evidence for the inverse relationship between the growth of industrial employment and the unemployment rate, further highlighting the importance of industrial sector expansion for improving labor market conditions in Brazil.

In recent studies, the methodology has been applied to assess the influence of high-tech and high-skilled jobs, as well as the creative industry. Goos, Konings and Vandeweyer (2018) conducted a study examining labor markets in 227 regions across Europe. Their findings reveal that additional employment in highly skilled occupations can generate up to 5 additional jobs in low-skill-intensive local services within the same region. However, the authors also observe persistent variations in the size of this multiplier across regions. They find that regions with higher levels of immigration, a larger number of less skilled workers, and lower GDP per capita tend to exhibit higher multipliers. This suggests that the characteristics of the region, such as labor market composition and economic development, play a role in shaping the multiplier effect.

A similar analysis was conducted by Lee and Clarke (2019) to examine the multiplier effect of high-tech industry jobs using UK labor market data. The study employed Moretti (2010)'s methodology and included a set of control variables. The findings revealed three main discoveries. Firstly, for every high-tech job created, approximately 0.7 jobs were generated in local services. Secondly, the study found that the growth of high-tech jobs led to a decrease in the average wage of low-skill local workers. However, the authors argued that this wage reduction was primarily driven by new entrants to the labor market, as employment of unskilled workers in the tradable sector remained unaffected.

Finally, Gutierrez-Posada et al. (2023) have recently applied this methodology to evaluate employment multipliers in the cultural and creative industries. The authors made adaptations to the methodology and constructed a 21-year panel dataset for cities in the UK. Their findings indicate that, on average, the addition of one creative job is associated with the creation of at least 1.9 new jobs in the tradable sector of each city. Furthermore, the creative industry accounts for over 16% of the growth in non-tradable employment in the analyzed sample, with more significant impacts observed in locations with larger creative clusters.

Indeed, these two distinct theoretical fields converge towards the central focus of this chapter. The literature on complexity has primarily centered around investigating the correlation between the Economic Complexity Index (ECI) and income distribution measures, while paying less attention to the exploration of regional inequalities. On the other hand, recent studies on regional employment multipliers, particularly concerning high-tech jobs, have evolved to inquire about the consequences of generating complex employment opportunities. Consequently, the connection between these areas becomes apparent. Thus, the primary objective of this chapter is to provide an answer to the following question: Do complex employment multipliers display heterogeneity based on the level of regional complexity?

The development process guided by path dependence, as explored in Chapter 2, suggests that our hypothesis leans towards a positive answer. Regions evolve economically in distinct ways, and as a result, we anticipate observing variations in employment multipliers. By studying these differences, we can gain insights into the aspects of regional inequality stemming from natural related diversification. It is therefore important to understand to what extent this dynamic is good or bad news, respectively, for the more and less complex Brazilian regions in terms of job creation.

3.2.2 Conceptual Framework Adapted for Economic Complexity Approach

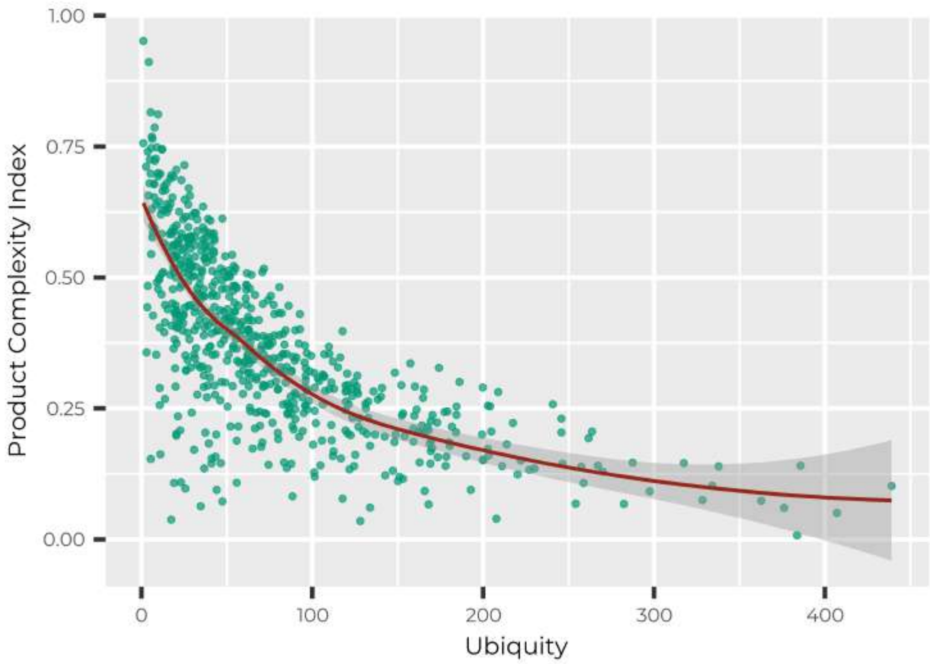
This analysis becomes possible by adapting the conceptual framework developed by Moretti (2010) and Moretti and Thulin (2013). As mentioned earlier, their framework is based on several key assumptions that enable the measurement of local employment multipliers. First, it considers each city as a competitive economy that produces two types of goods and services: tradables and non-tradables. Tradables have prices determined at the national or international level, outside the control of the cities, while non-tradables have locally determined prices.

Moreover, the framework assumes that labor is perfectly mobile across sectors within cities. This assumption ensures that the marginal product and the marginal wage are equalized within cities, leading to efficient allocation of labor resources. Additionally, the utility of workers is influenced by local net wages, the cost of living, and individual preferences for specific locations. The extent of idiosyncratic location preferences affects the geographic mobility of labor, with weaker preferences leading to greater labor mobility and higher elasticity of labor supply. It is also assumed that the housing supply curve is upward sloping, with the slope being influenced by geographical factors and land use regulations. These assumptions collectively form the foundation for estimating local employment multipliers.

The adaptation of these assumptions with the concepts of economic complexity is facilitated by the nature of the main indicators used in this approach. As discussed in section 2.3.1, the Product Complexity Index (PCI) takes into account the concepts of diversification and ubiquity. Ubiquity refers to the widespread production of a good across different regions, indicating lower complexity. On the other hand, goods that are produced by fewer regions competitively are considered more complex. This notion of ubiquity allows us to classify the economy into the production of complex and non-complex goods and services, as well as tradable and non-tradable sectors.

The greater the presence of a sector in local economies (high ubiquity), the more influential local dynamics are in determining its price. Conversely, sectors with a smaller presence in local productive structures (low ubiquity) are less susceptible to local factors in determining their price. As a result, highly ubiquitous activities such as the sale of beverages, food, construction materials, bakery products, cleaning services, and accommodation services are considered less complex. These activities compete with other local actors, and their prices are primarily defined locally. On the other hand, activities with lower ubiquity, such as manufacturing, financial and banking services, and information intelligence services, are more complex and compete at the national or international level. Therefore, their prices are largely determined by factors beyond the local context. The graph below illustrates the comparison between the levels of complexity and ubiquity for a range of productive activities in Brazil.

Figure 5 – Product Complexity Index (PCI) and Ubiquity - Brazilian Economic Activities



Source: own elaboration.

Therefore, our conceptual framework begins by distinguishing between complex and non-complex goods and services. Each micro-region in Brazil can be seen as a competitive economy that allocates its workforce between the production of complex and non-complex sectors. The prices of complex goods and services are determined outside the local dynamics, while non-complex sectors are influenced by local factors. Finally, the hypotheses related to labor force mobility and the upward sloping labor and housing supply curves still apply to the analysis of local employment dynamics within the region.

Our research focuses on examining the impact of shocks on labor demand in both the complex and non-complex sectors of the economy, similar to the approach taken by Moretti (2010) in studying the tradable sector. Specifically, we are interested in understanding the effects of permanent growth in these sectors, whether it be through the attraction of new industries or exogenous increases in labor productivity within existing industries. These shocks not only directly affect employment in the respective sectors but also have indirect effects on the rest of the economy. It is important to note that such shocks may also have implications for the general price equilibrium, as they are likely to result in increased wages for workers and higher housing costs, unless labor and housing supplies are infinitely elastic at the local level.

Moretti (2010) focuses on measuring these two indirect effects. Firstly, there is the multiplier effect on the non-tradable sector, which is a result of the increase in aggregate income within cities. This increase in employment and local wages leads to higher demand in the non-tradable sector. Secondly, there is the multiplier effect on the remaining tradable sector, which is influenced in various ways. Employment in this sector may decrease due to the rising labor costs and decreased competitiveness. Conversely, it may increase if there is a concentration of intermediate tradable goods production locally or due to agglomeration economies. Therefore, our study shares a similar objective, as we seek to examine these dynamics and their implications in the context of complex and non-complex sectors.

We will utilize this adapted conceptual framework to examine the impacts of shocks on both the complex and non-complex sectors of the Brazilian economy. Our analysis will encompass all potential relationships between these sectors and will further differentiate regions based on their level of complexity, as studied in Chapter 2. Through this investigation, we aim to validate the following hypotheses.

- *Hypothesis 1*: Complex sector employment multipliers are greater.
- *Hypothesis 2*: Multipliers are heterogeneous between regions with different levels of economic complexity.
- *Hypothesis 3*: Job creation by the complex sector is more effective in regions with highly complex production structures.

Hypothesis 1 is based on the premise that the attraction of complex (less common) jobs, such as manufacturing activities, leads to an increased demand for less complex (more common) activities, such as basic services. This relationship is analogous to the one observed between tradables and non-tradables. Conversely, the opposite reasoning lacks the necessary transmission channels to generate significant job creation. Hypothesis 3 follows as a consequence of Hypothesis 2, suggesting that the complex sector has varying impacts on regions based on their existing level of complexity. This is rooted in the understanding that more complex economies are also institutionally more developed regions, characterized by lower wage inequality between sectors and higher competitiveness. As a result, the labor supply in these regions is more sensitive to changes caused by permanent increases in employment within the complex sector. On the other hand, less complex regions, with different characteristics, are not affected in the same manner by the complex sector.

The analysis of these hypotheses is unprecedented and distinct from other studies that have examined the duality between tradable and non-tradable sectors. While other adaptations have been made to understand employment multipliers across sectors such as the public sector, high-tech industries, and creative and cultural industries, this complexity approach study offers a unique perspective. For example, it allows evaluating not only the multiplier of manufacturing activities but also more complex services. That is, the differentiation between complex and non-complex sectors encompasses the multipliers of the production of goods and services characterized by sophisticated skills and capabilities that would be left out when analyzing, for instance, only the tradable sector or high-tech industries. As a result, it captures spillover effects in the economy that were not captured in previous analyses with different segmentation.

3.3 Data and Method

3.3.1 Data

For the analysis in this chapter, the primary database will consist of employment data in economic activities across micro-regions in Brazil for the years 2009, 2014, and 2019. These data are sourced from the Annual Social Information Report (RAIS), which is linked to the Brazilian Ministry of Labor and Employment. It is important to note that the RAIS database contains administrative records of all formal establishments in the Brazilian labor market, making it a crucial data source for similar methodological analyses conducted in previous studies on Brazilian regions (MACEDO; MONASTERIO, 2016; LOYO; MOISÉS; MENDES, 2018; ROCHA; ARAÚJO, 2021).

Similar to Rocha and Araújo (2021), the geographic unit used in this study will be the micro-regions. Economic activities will be classified based on the 6-digit class of the National Classification of Economic Activities (CNAE) provided by the Brazilian Institute of Geography and Statistics (IBGE). The choice of micro-regions and the specific classification system aligns with the previous chapter's justifications and facilitates the use of consistent indicators for both analyses. By using the same database and indicators, comparability and continuity between the chapters are ensured.

Additionally, although data segmented according to this classification are available from 2006 to 2021, we have chosen to maintain the same time period as in Chapter 2, which includes the year intervals 2009-2014 and 2014-2019. This choice aligns with the literature on multipliers, ranging from classic articles (MORETTI, 2010; MORETTI; THULIN, 2013) to more recent studies (DIJK, 2017; MACEDO; MONASTERIO, 2016; ROCHA; ARAÚJO, 2021), which have also used three time points and two intervals for their analyses.

This classification enables us to create a database that includes 558 Brazilian micro-regions and 670 productive activities, resulting in a total of 373,860 observations per year. Since we are using three time points, the database consists of a total of 1,121,580 observations. However, as explained in the next subsection, our econometric estimation strategy is aggregated at the micro-region level and focuses on the variation in employment between periods. Consequently, the final dataset used for the estimation comprises 1,116 observations. The subsequent subsection provides further information on the specific procedures employed in our analysis.

3.3.2 Econometric Specifications

The econometric specification employed to measure the multipliers will be adapted from the approach used by Moretti and Thulin (2013). Their methodology allows for the identification of the indirect effects resulting from a permanent increase in the tradable sector. However, in line with our conceptual framework, we will adapt this approach to examine the indirect effects of exogenous employment shocks in both the complex and non-complex sectors. To accomplish this, we will estimate four equations that capture all possible relationships between these two sectors, which together constitute the economy.

These equations are listed below. The first two, 3.3 and 3.4, calculate the multipliers resulting from a shock in the complex sector. In 3.3, this effect is verified on the non-complex sector of the economy. In 3.4, a part of the complex sector is randomly selected to check the employment multiplier over the rest of the complex sectors. The logic is the same for the non-complex sector in equations 3.5 and 3.6. Respectively, the effect on the complex sector and of a portion of the non-complex sector on the rest of the same sector is calculated.

$$E^{NC}_{m,t} - E^{NC}_{m,t-5} = \beta_0 + \beta_1(E^{C}_{m,t} - E^{C}_{m,t-5}) + \beta_2 Time + \varepsilon_{m,t} \tag{3.3}$$

$$E^{C1}_{m,t} - E^{C1}_{m,t-5} = \beta_0 + \beta_1(E^{C2}_{m,t} - E^{C2}_{m,t-5}) + \beta_2 Time + \varepsilon_{m,t} \tag{3.4}$$

$$E^{C}_{m,t} - E^{C}_{m,t-5} = \beta_0 + \beta_1(E^{NC}_{m,t} - E^{NC}_{m,t-5}) + \beta_2 Time + \varepsilon_{m,t} \tag{3.5}$$

$$E^{NC1}_{m,t} - E^{NC1}_{m,t-5} = \beta_0 + \beta_1(E^{NC2}_{m,t} - E^{NC2}_{m,t-5}) + \beta_2 Time + \varepsilon_{m,t} \tag{3.6}$$

Therefore, $E^{NC}_{m,t}$ and $E^{C}_{m,t}$ represent the amount of employment, respectively, in the non-complex and complex sectors in micro-region m and in period t. $E^{C1}_{m,t}$ reflects employment in a randomly selected portion of the complex sector and $E^{C2}_{m,t}$ the amount of employment in the rest of the sector. The same is represented for $E^{NC1}_{m,t}$ and $E^{NC2}_{m,t}$ for non-complex sector parts. The $Time$ variable is a dummy that takes the value 1 referring to the last period (2014-2019). This strategy is adopted to control for possible national shocks in employment in the sector that is the dependent variable. Finally, $\varepsilon_{m,t}$ is the error term. In all equations, the regional employment multiplier is hypothetically represented by β_1.

However, the OLS estimations of these models are likely to be inconsistent. As summarized by Dijk (2015), this is because β_1 is capturing three types of effects. First, it captures the causal effect of job growth in one sector on the other, which is the effect we want to measure. Second, there is likely to be an endogeneity problem, such as when an increase in jobs in the non-complex sector affects the number of jobs in the complex sector (equation 3.3) or vice versa (equation 3.5). Third, there may be inconsistencies due to omitted variables, such as changes caused by local public services that influence employment in both sectors.

To address these problems, Moretti and Thulin (2013) propose using an instrumental variable estimation with a shift-share instrument (BARTIK, 1991). The shift-share analysis decomposes employment growth into three distinct effects: growth resulting from the increase in total national employment (national), growth due to the composition of local productive structures (structural), and growth resulting from the performance of these sectors locally compared to the performance of the same sectors in the overall economy (differential). The strategy employed by Moretti and Thulin (2013) is to isolate potentially exogenous changes in job demand by calculating the structural growth component. In this case, the instrumental variable aims to isolate the variation in employment in the tradable sector that is due to national changes, separate from the variation that is due to local changes. For our purposes, we have adapted the calculation of this instrument to suit our analysis:

$$IV_1 = \sum_j E^{\phi}_{m,j,t-5}(ln(E^{\phi}_{j,t} - E^{\phi}_{m,j,t}) - ln(E^{\phi}_{j,t-5} - E^{\phi}_{m,j,t-5})) \tag{3.7}$$

Where $\phi = \{$C (equation 3.3), C2 (3.4), NC (3.5), NC2 (3.6)$\}$.

This instrument, represented by equation 3.7, includes the national share and the sector-specific shares but excludes regional variation. Unlike Moretti (2010), where the instrument did not exclude the variation of the city itself, this adaptation proposed by Moretti and Thulin (2013) addresses the potential violation of the exogeneity assumption. By excluding the variation of the reference micro-region, the instrument isolates changes in employment in industry j of micro-region m that arise from national variations in industry j. However, the impact of these changes differs across micro-regions due to their unique composition in the base year ($E^{\phi}_{m,j,t-5}$). According to Moretti and Thulin (2013), the instrument captures exogenous changes in local labor demand, as national changes do not reflect local economic dynamics.

In addition, we will utilize another type of instrument to assess potential changes in the multiplier and enhance the robustness of our results. The instrument proposed in equation 3.7, as mentioned earlier, is based on the productive structure composition of the micro-regions in the base year, thereby not capturing the effects of structural changes within each local economy over the 5-year interval. To address this limitation, our complementary approach involves using an instrument that isolates changes resulting from national variations and local variations, taking into account the portfolio of economies in the final period ($E^{\phi}_{m,j,t}$) rather than the initial period. Consequently, changes in the overall economy will continue to manifest differently across micro-regions, but now considering the structural changes observed during the period. This adaptation is inspired by Stilwell (1969)'s proposal, which modifies the shift-share method to calculate the expected net change in employment in a given region based on its final industrial structure. The author's approach is based on the criticism that the conventional method does not capture the diversification that occurred during the studied period. The calculation of this shift-share instrument is shown below:

$$IV_2 = \sum_j E^{\phi}_{m,j,t}(ln(E^{\phi}_{j,t-5} - E^{\phi}_{m,j,t-5}) - ln(E^{\phi}_{j,t} - E^{\phi}_{m,j,t})) \tag{3.8}$$

Still, two important points regarding the estimation strategy should be addressed. Firstly, it is worth noting that most of the studies in this literature do not incorporate control variables in their estimations, although there are exceptions (FAGGIO; OVERMAN, 2014; DIJK, 2017; WANG; CHANDA, 2018; LEE; CLARKE, 2019). In these cases, the variables are typically used to control for city or regional size, the skill level of the workforce or inhabitants, and the unemployment rate. In our analysis, we will replicate the estimations with the inclusion of control variables to assess the robustness of our results. Specifically, we will control for the population size of each micro-region using data from IBGE. Additionally, we will consider the share of employment occupied by individuals with at least an incomplete undergraduate degree as a measure of the region's labor market qualification. The average salary of the micro-region will be used to control for productivity, and the local relatedness average will be employed to account for the proximity of the micro-region to other sectors, which is expected to influence the attraction of new jobs. With the exception of population data obtained from IBGE, the variables will be constructed based on data provided by RAIS. It is important to note that unemployment rate at the micro-region level is not available. The estimations including these control variables will be available in the annexes.

Second, a different estimation strategy is employed for equations 3.4 and 3.6. In the literature that examines the effect of a portion of the tradable sector on the rest of the same sector (MORETTI, 2010; MORETTI; THULIN, 2013), there is no specific guidance on how the samples are selected. The authors only mention that a part of the sector is randomly chosen. However, when conducting a single estimation and selecting only one sample, the resulting multiplier is solely determined by that particular sample. As a result, it is not possible to assess the sensitivity of the multiplier to the sample selection process. In other words, the value of the multiplier may significantly differ if a different sample were randomly selected. To address this concern, we will conduct a *bootstrap analysis*[1]. By randomly selecting multiple samples, we can determine the variability of the resulting multipliers and their trends. This approach allows for a more robust estimate that is not contingent on the estimation based on a single sample.

Finally, it is worth emphasizing that the estimates will be conducted for all micro-regions as well as for sample groups based on their level of complexity. The objective of examining the multipliers for regions according to their complexity is to explore one

[1] See Efron and Tibshirani (1993).

aspect of regional inequality resulting from diversification into sectors with varying levels of complexity. The classification of the regions' complexity level and the criteria for categorizing complex and non-complex sectors will be discussed in the subsequent section along with the descriptive analysis of the data.

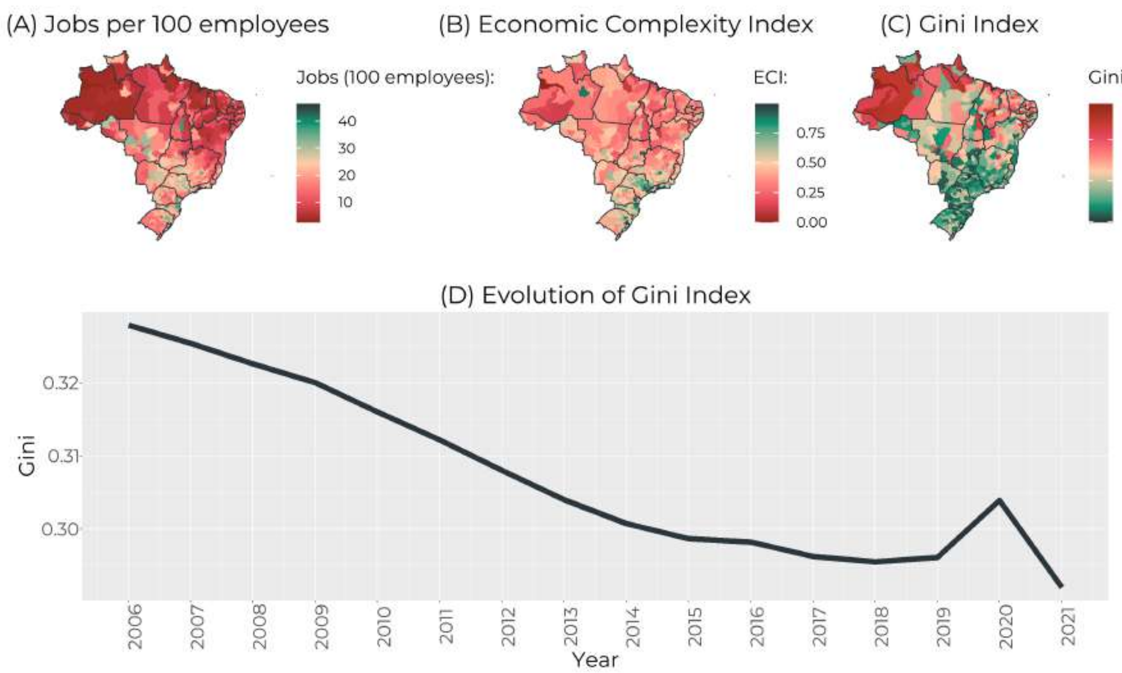
3.4 Descriptive Analysis

3.4.1 Regional Inequality in Brazil

Brazil is divided into 558 micro-regions, which represent the country's vast geographical and demographic diversity. However, the analysis pattern remains consistent regardless of the variable being examined. In terms of employment, the micro-regions comprising the three main capitals of the Southeast region (São Paulo, Rio de Janeiro, and Belo Horizonte) account for approximately 22% of total employment during the study period. In relative terms, these micro-regions, along with others, rank at the top in terms of average employment per inhabitant. Figure 6A illustrates the number of jobs per 100 inhabitants in each micro-region, indicating a concentration of employment in large urban centers. Micro-regions with state capitals and other major urban centers are prominent, particularly in the states of São Paulo, Paraná, and Santa Catarina. Conversely, the North and Northeast regions are characterized by lower employment rates per population, indicated by the red color on the map.

The distribution of the complexity index aligns with the employment distribution discussed earlier. There is a strong correlation of over 80% between economic complexity and the employment rate per inhabitant. Figure 6B illustrates the concentration of complexity in major urban centers, confirming the structural difference in complexity across regions (BALLAND et al., 2020). This regional inequality in terms of complexity is significant because it poses limitations to sophisticated diversification, as discussed in Chapter 2. It is expected that this disparity in complexity will impact the magnitude of employment multipliers, and this hypothesis will be tested through the econometric estimates.

Figure 6 – Inequality, Employment and Complexity in Brazilian Micro-regions



Source: own elaboration.

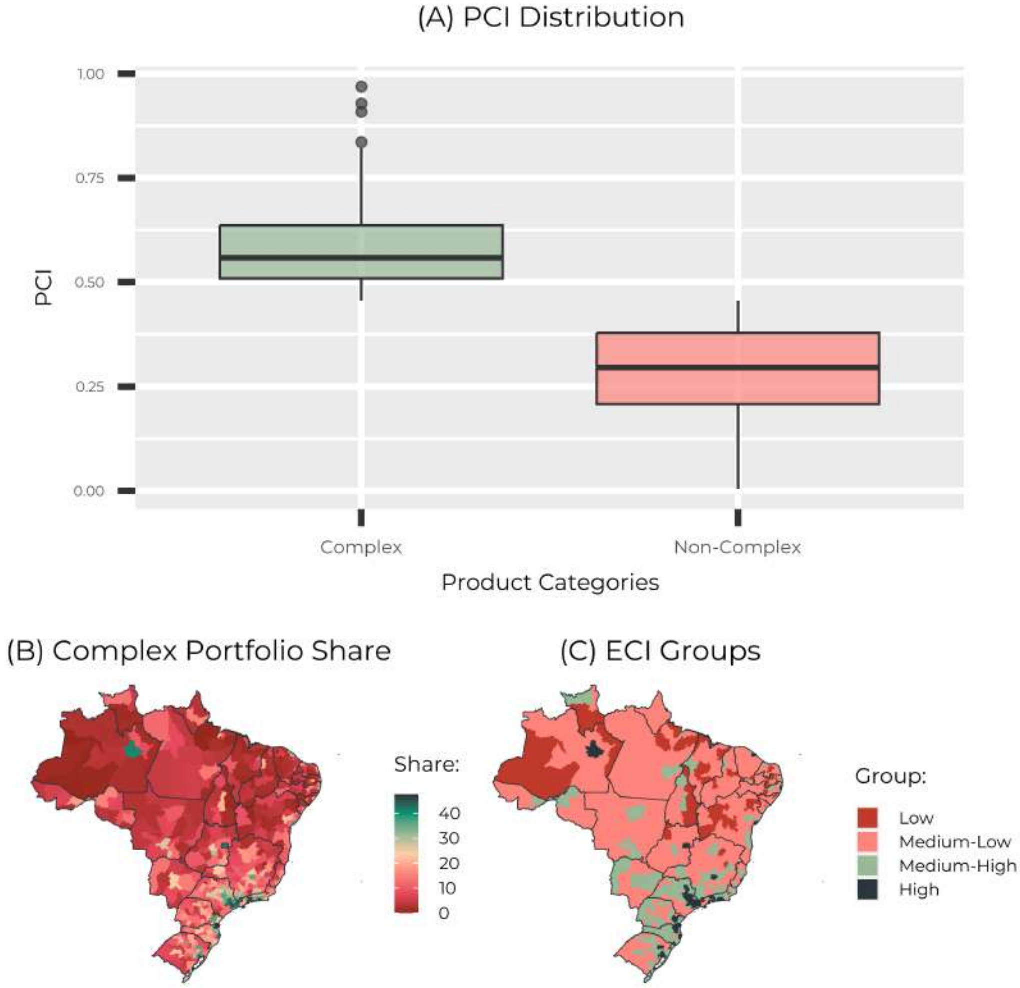
To assess regional inequality in terms of employment in economic activities, we employed the Gini index, following the approach used by Moretti and Thulin (2013). The Gini index was calculated based on the average wage for productive activities in each micro-region, using data from RAIS. This measure provides insights into wage inequality among activities and micro-regions. Figure 6C displays the Gini index and highlights the concentration of the greatest inequalities in the interior of the Northeast and North regions. In the Southeast and South regions, the highest inequality is observed in the northern part of Minas Gerais. The data corroborates the same pattern observed in other indicators. For instance, there is a correlation of over 70% between wage inequality (Gini index), per capita employment rate, and economic complexity. Notably, Figure 6D depicts the average Gini index for micro-regions from 2016 to 2021, showing a downward trend over the period. The 2020 peak is associated with the crisis resulting from the pandemic, in which inequalities worsened.

3.4.2 Complexity Classification

The classification of complex and non-complex sectors in this study differs from the classification of tradables and non-tradables. Instead, we use an indicator that reflects the complexity of each sector, specifically based on the Product Complexity Index (PCI). As there are no studies that directly relate the complexity approach to the calculation of multipliers, we have established criteria to differentiate economic activities based on their

PCI values. We adopted two strategies to ensure robustness. The first strategy is more aggressive, classifying activities in the last tertile of the PCI as "complex" and those in the first and second tertiles as "non-complex". Additionally, we conducted robustness tests considering the PCI value itself, rather than its distribution. The second strategy categorizes activities as "complex" when they have a positive PCI and as "non-complex" when they have a negative PCI, before normalizing the indicator between 0 and 1. This classification includes many more sectors as complex compared to the previous categorization. Estimates with the second classification are available in Annex B.

Figure 7 – Complexity Classification - Sectors[1] and Micro-regions[2]

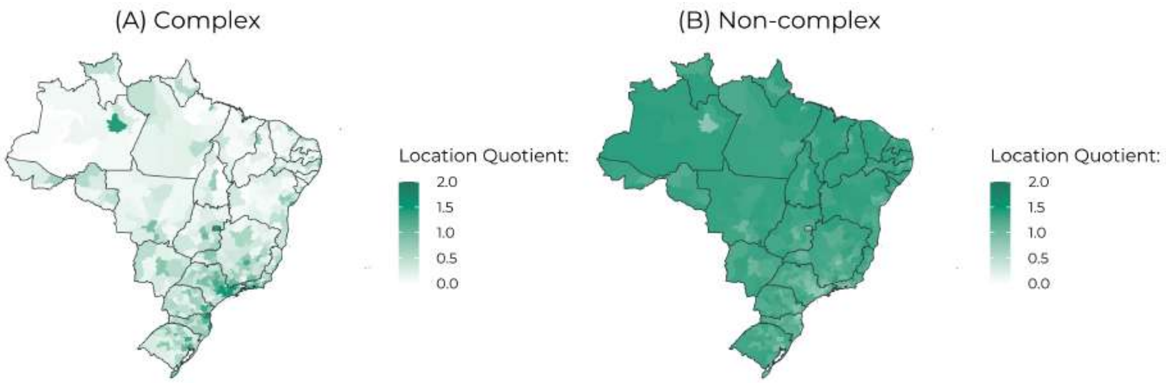


Source: own elaboration.

[1]Complex sectors are those falling within the third tertile of PCI values, while non-complex sectors are in the first and second tertiles. [2]The micro-regionswere categorized based on ECI values as follows: Low $(0,00 \leq ECI \leq 0,25)$, Medium-Low $(0,25 < ECI \leq 0,50)$, Medium-High $(0,50 < ECI \leq 0,75)$, High $(0,75 < ECI \leq 1,00)$.

Figure 7A displays the distribution of the Product Complexity Index (PCI) within each category. The complex sector has a more skewed distribution than the non-complex one. In this first group, the distribution is more concentrated on the lower limit, the median is around 0.556 and there are outliers with the PCI close to 1. Among the non-complex ones, the PCI is better distributed, outliers are not perceptible and the median is close to 0.295. Figure 7B demonstrates the size of the complex sector's participation in local economies. Obviously, the highest participation rates are in the most complex regional economies. Finally, Figure 7C rescues the groups of micro-regions according to complexity. These 4 groups will also be used to differentiate the multiplier estimates. As a reminder, the classification of microregions is based on the resulting ECI value: Low (ECI between 0 and 0.25), Medium-Low (0.25 and 0.50), Medium-High (0.50 and 0.75) and High (0.75 and 1.00).

Figure 8 – Location Quotient of Complex and Non-Complex Sectors



Source: own elaboration.

Building on the analysis conducted by Rocha and Araújo (2021), we calculated the location quotient for both the complex sector (Figure 8A) and the non-complex sector (Figure 8B) to evaluate their distribution and specialization within the region. The data reveals that employment in the complex sector is more concentrated, especially in the Southeast and South regions. In contrast, employment in the non-complex sector shows a more even distribution across the entire territory, with a significant presence in all regions of the country. This distribution pattern supports the notion that non-complex activities are influenced more by local dynamics and driven by local consumption. After discussing the sector classification, we will now present the results of the econometric tests.

3.5 Econometric Results

This chapter aimed to estimate four main regressions (equations 3.3 to 3.6) that capture all possible relationships between complex and non-complex sectors. Additionally, to examine potential heterogeneity across regions, each regression was estimated for different levels of complexity. Following the approach of previous literature (MORETTI; THULIN, 2013; MACEDO; MONASTERIO, 2016), we will provide a summary of the results to facilitate understanding. Tables 9 and 10 present the multipliers for each model specification and level of complexity. Detailed results for each model can be found in the Annex B.

Therefore, Table 9 ppresents the results of employment multipliers in complex sectors over employment in non-complex sectors. Columns (1) and (2) show the OLS estimates without and with controls for observable characteristics that vary over time in the micro-regions. The remaining columns follow the same pattern but include instrumental variable estimation to address potential endogeneity issues. Columns (3) and (4) utilize the instrumental variable proposed by Moretti and Thulin (2013), specified in equation 3.7. Columns (5) and (6) present the results using the instrumental variable proposed in this study, taking into account the structural changes in local economies over the period, as represented by equation 3.8.

Table 9 – Complex Employment Multipliers over Non-complex Employment

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
General Model	1.76*** (0.35)	1.71*** (0.37)	2.25*** (0.33)	2.36*** (0.32)	1.89*** (0.32)	1.87*** (0.33)
Low	−2.26 (2.01)	−1.90 (1.85)	1.95 (4.66)	1.33 (4.61)	−2.71 (2.43)	−2.45 (1.93)
Medium-Low	1.74** (0.78)	0.83 (0.61)	22.17 (20.85)	19.53 (18.96)	2.20*** (0.79)	1.15* (0.60)
Medium-High	1.95*** (0.29)	1.86*** (0.32)	3.25*** (0.32)	3.50*** (0.44)	2.05*** (0.29)	1.98*** (0.34)
High	1.56*** (0.39)	1.51*** (0.37)	2.18*** (0.39)	2.39*** (0.45)	1.71*** (0.36)	1.72*** (0.34)
Controls	No	Yes	No	Yes	No	Yes

Signif.: *p<0.1; **p<0.05; ***p<0.01
Robust standard-errors (clustered at the micro-region) are in parentheses.

As anticipated, the complex sector has a significant impact on the rest of the economy in most cases. On average, the results indicate that one new job in the complex sector leads to the generation of 1.76 to 2.36 additional jobs in the non-complex sector. Furthermore, when controlling for factors such as human capital, proximity, average salary, and the size of the micro-region, the magnitude of the multipliers remains consistent. However, it is important to note that the choice of instrument affects the magnitude of the multiplier. The use of an instrument that considers the variation in the composition of local employment in the final period reduces the multiplier by approximately 0.3. Nevertheless, the overall findings are statistically and economically significant in all cases.

The analysis of the multipliers by complexity groups reveals a more diverse scenario. The results indicate that the most significant effects are observed in regions that are already complex. These regions demonstrate the capability of their complex sectors to have a substantial impact on the rest of the economy when experiencing a permanent increase in employment. However, in less complex regions, the existing local complex sector appears to be insufficient to exert a significant influence on the rest of the economy. Interestingly, among the Medium-Low regions, the results show a notable multiplier effect when considering the effect of structural change in the instrumental variable estimation. This effect is not observed when using the classic instrument. One possible explanation for this finding is that the structural change in these regions has made their economies slightly more complex, enabling a minimum level of complexity that facilitates an increase in employment in the rest of the economy. On the other hand, the Medium-High and High regions consistently demonstrate significant multipliers at 1% in all estimations. For the Medium-High regions, an increase of one job in the complex sector leads to the generation of 3.5 additional jobs in the non-complex sector. Similarly, the High complexity regions experience an increase in employment ranging from 1.51 to 2.39 jobs in the non-complex sector for each additional job in the complex sector.

Table 10 – Non-complex Employment Multipliers over Complex Employment

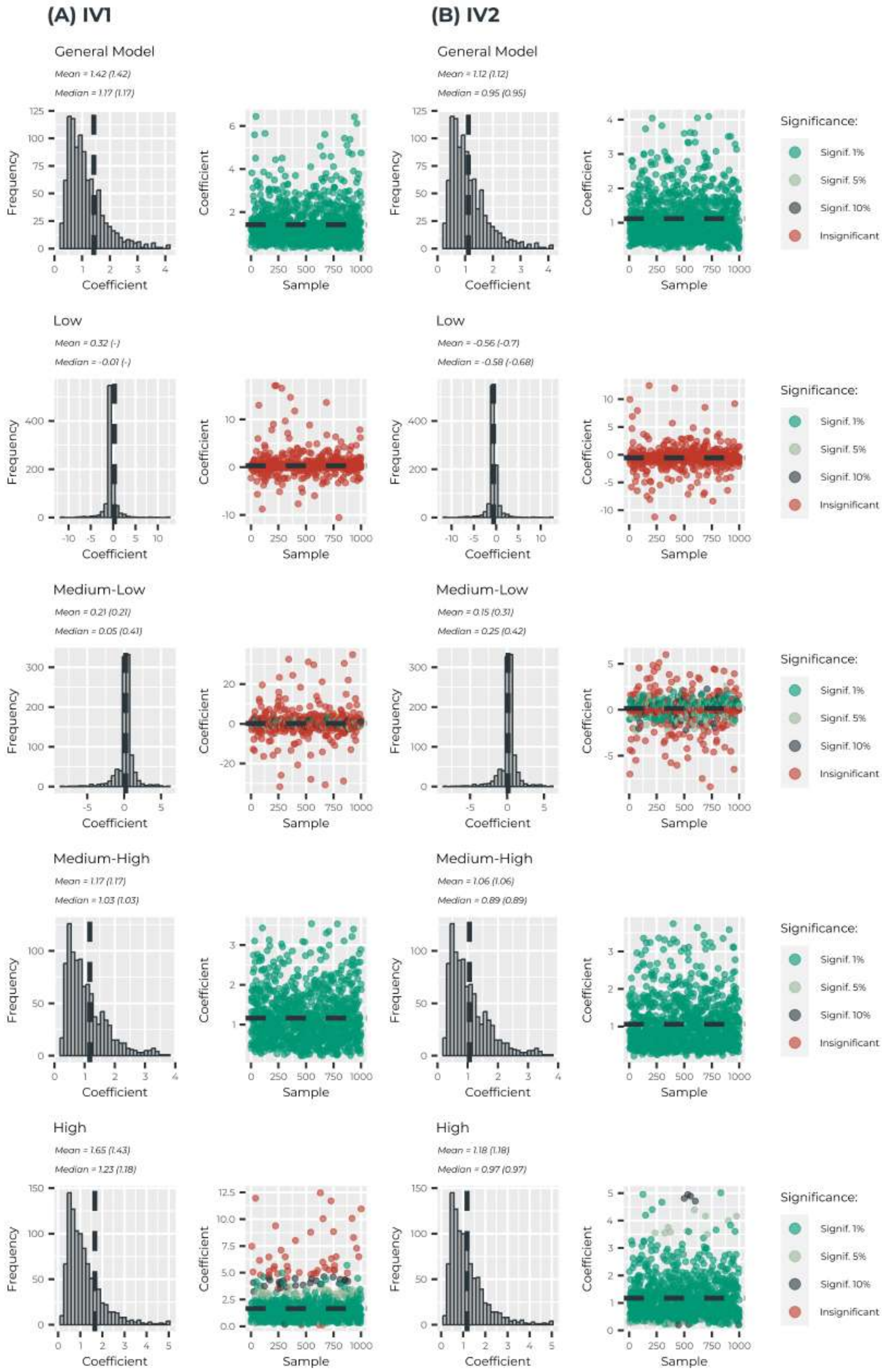
	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
General Model	0.39*** (0.05)	0.38*** (0.04)	0.45*** (0.06)	0.43*** (0.05)	0.39*** (0.05)	0.37*** (0.05)
Low	−0.01* (0.00)	−0.01* (0.00)	−0.01** (0.00)	0.00 (0.01)	−0.01** (0.00)	−0.01*** (0.00)
Medium-Low	0.04*** (0.01)	0.02* (0.01)	0.05** (0.02)	0.05 (0.03)	0.04*** (0.01)	0.02** (0.01)
Medium-High	0.22*** (0.04)	0.18*** (0.03)	0.28*** (0.04)	0.28*** (0.04)	0.24*** (0.03)	0.19*** (0.04)
High	0.40*** (0.06)	0.38*** (0.06)	0.47*** (0.07)	0.43*** (0.06)	0.41*** (0.07)	0.36*** (0.06)
Controls	No	Yes	No	Yes	No	Yes

Signif.: *p<0.1; **p<0.05; ***p<0.01
Robust standard-errors (clustered at the micro-region) are in parentheses.

Table 10, similar to Table 9, presents the estimates of the non-complex employment multipliers over employment in the complex sector. These estimates provide insights into the inverse effect, correcting for endogeneity using instrumental variables. As expected, the results demonstrate smaller magnitudes of multipliers. Job growth in highly ubiquitous and less complex sectors does not have a substantial impact on the more sophisticated sectors of the economy, in comparison to the reverse relationship. On average, one additional job in the non-complex sector results in the generation of 0.37 to 0.45 additional jobs in the complex sector. These findings highlight the asymmetric relationship between the complex and non-complex sectors, with the complex sector exerting a stronger influence on job creation in the non-complex sector compared to the reverse relationship.

Among the complexity groups, the regions with higher complexity levels continue to exhibit the highest multipliers. It is evident that the absence of a robust complex sector in low and medium-low complexity economies limits the potential impact of permanent increases in the non-complex sector. In these economies, the multipliers are close to zero or even negative, suggesting a crowding-out effect in the complex sector. Conversely, highly complex micro-regions show significant impacts, with the multipliers approaching the overall effect. In these regions, an increase of one job in the non-complex sector generates between 0.36 and 0.47 new jobs in the complex sector. These findings highlight the importance of a strong and developed complex sector in driving employment generation.

Figure 9 – Complex-Complex Multiplier



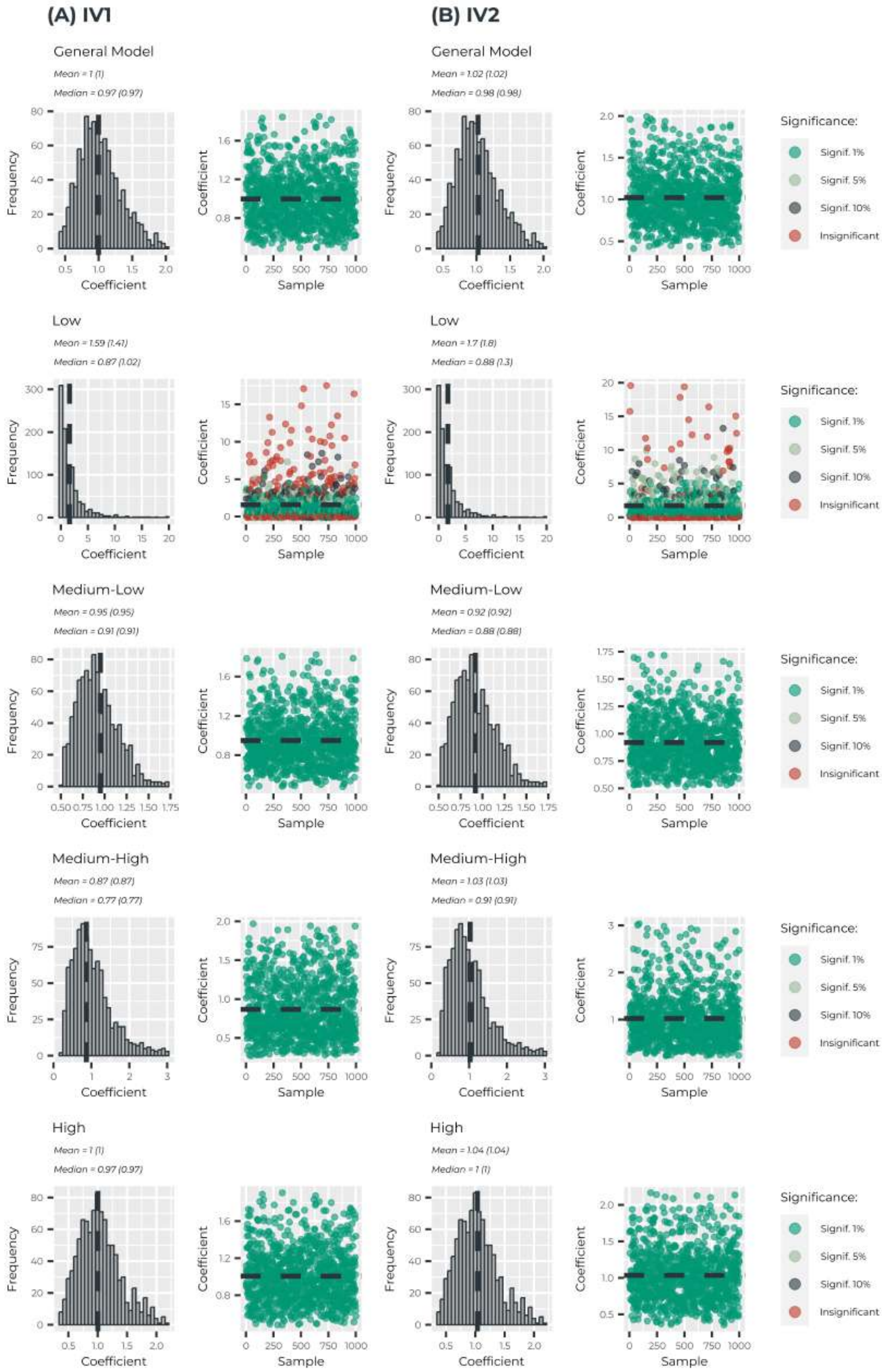
Source: own elaboration.

Finally, Figures 9 and 10 provide a summary of the remaining estimates, focusing on the influence of a portion of economic activity within the complex and non-complex sectors, respectively, on the rest of the activities in the same sector. Since the literature does not specify the method of selecting the sample for this analysis, we employed a bootstrap approach, randomly selecting parts of the reference sector 1,000 times. For clarity, the bootstrap estimations presented in the figures only include the instrumental variable specification without control variables. As observed in the results presented in Tables 9 and 10, the inclusion of control variables does not significantly alter the magnitude and significance of the multipliers.

Therefore, Figure 9 displays the results for the estimates of equation 3.4, with two columns representing different instruments. Column (A) shows the estimates using the conventional shift-share instrument, while column (B) presents the estimates with the adapted shift-share instrument. Each row corresponds to a type of region or the general model, and for each column, two graphs are provided. The first graph is a histogram plot depicting the distribution of estimated multipliers across the replicates, and the second graph shows the multiplier values for each sample from 1 to 1,000, indicating whether they are statistically significant. Additionally, the mean and median values of the resulting multipliers are provided for each estimate (A and B). The values in parentheses represent descriptive statistics for the significant multipliers up to 10% significance level. Figure 10 follows the same format but presents the results for equation 3.6.

Figure 9 illustrates that the average multiplier for the rest of the complex sector, resulting from an increase of 1 job in a portion of the complex sector, ranges from 1.12 to 1.42. However, the level of complexity in the micro-regions affects the magnitude of this multiplier. In less complex regions, the multiplier is less robust, with only a few estimates being statistically significant, particularly in low complexity micro-regions. For Medium-Low, the significant multiplier estimates are 0.21 (A) and 0.31 (B). As the level of complexity increases, the magnitude of the multiplier also increases. High complexity regions exhibit stronger performance compared to the general case. With the conventional shift-share instrument, an increase of 1 job in a portion of the complex sector generates, on average, 1.65 jobs in the rest of the sector. These findings align with the earlier results, indicating the limited efficiency of the complex sector in less complex regions.

Figure 10 – Non-Complex-Non-Complex Multiplier



Source: own elaboration.

However, the evaluation of non-complex multipliers reveals a slightly different scenario. Figure 10 reveals that, on average, the multiplier for the rest of the non-complex sector resulting from an increase of 1 job in a portion of the non-complex sector is 1-to-1. In other words, the increase of one job in the non-complex sector generates one job in the rest of the sector. Interestingly, Low micro-regions stand out in this case. The significant multipliers for the non-complex sector in these regions are 1.41 (A) and 1.8 (B). This indicates that the impact of job creation in the non-complex sector is more pronounced in low complexity micro-regions. Additionally, when considering the structural change of these regions, there is an observed increase of 0.4 in the average multipliers for the non-complex sector. On the other hand, the other micro-regions exhibit multipliers similar to the general average in both types of estimates.

Table 11 – Multipliers Summary Table

	Multiplier:	
	Complex Employment Multiplier over Non-complex Employment	
	IV_1 (1)	IV_2 (2)
General Model	2.25***	1.89***
Low	1.95	−2.71
Medium-Low	22.17	2.20***
Medium-High	3.25***	2.05***
High	2.18***	1.71***
	Non-complex Employment Multiplier over Complex Employment	
	IV_1 (1)	IV_2 (2)
General Model	0.45***	.39***
Low	-0.01**	-0.01**
Medium-Low	0.05**	0.04***
Medium-High	0.28***	0.24***
High	0.47***	0.41***
	Complex Employment Multiplier over Complex Employment[b]	
	IV_1 (1)	IV_2 (2)
General Model	1.42	1.12
Low	-	−0.7
Medium-Low	0.21	0.31
Medium-High	1.17	1.06
High	1.43	1.18
	Non-complex Employment Multiplier over Non-complex Employment[b]	
	IV_1 (1)	IV_2 (2)
General Model	1.00	1.02
Low	1.41	1.80
Medium-Low	0.95	0.92
Medium-High	0.87	1.03
High	1.00	1.04

[b] As the estimates were via bootstrap, we report the mean value of the multipliers that were significant up to 10%.
Signif.: *p<0.1; **p<0.05; ***p<0.01

The results of the econometric tests are summarized in Table 11. As the multipliers do not vary considerably with the inclusion of control variables, we report the two main specifications: $IV_1(1)$ refers to the regression only with the conventional shift-share instrument, and $IV_2(2)$ with the adapted shift-share instrument. In Table 11, we observe that the complex sector demonstrates the highest multipliers, signifying its substantial impact on local employment. However, the extent of this influence varies significantly among micro-regions, depending on their level of complexity. In regions characterized by lower complexity, the complex sector exhibits limited effects, not only on itself but also on the non-complex sector. In contrast, complex regions experience a pronounced influence from the complex sector on the labor market, presenting the most potent multipliers.

3.6 Concluding Remarks

Covering gaps in the literature on complexity and regional inequalities, this chapter adapts the methodology of local employment multipliers (MORETTI, 2010; MORETTI; THULIN, 2013) to assess the regional multipliers of complex and non-complex sectors. The concern of the most recent literature to evaluate high-tech job multipliers (LEE; CLARKE, 2019) was one of the justifications for understanding this conceptual framework under the complexity approach. In addition, the relationship between these sectors and their magnitude were taken into account to verify how good or how bad it is to have, respectively, a productive structure towards more complex or less complex sectors. This intention resides in the scarcity of literature in evaluating possible implications of the uneven regional development to which the regions are submitted in light of complexity (PINHEIRO et al., 2022).

The descriptive analysis of data from the formal labor market in Brazilian micro-regions reveals important insights. Brazil is a highly unequal country, in which the volume of jobs, economic complexity and lower wage inequality are concentrated in the same regions. The micro-regions made up of state capitals and a few interior regions located in the Southeast and South regions stand out. Regional inequality, measured by the average wage between economic activities in the micro-regions, is considerably higher in the North and Northeast. Nevertheless, there is a decreasing trend in average wage inequality over time. In addition, obviously, when classifying complex and non-complex sectors, the same pattern of concentration is observed for the productive structures more focused on complex activities. However, the analysis of the employment share of micro-regions in these sectors in national employment reveals a different scenario. As expected, the greater participation of the non-complex sector is more widespread, since these are more ubiquitous activities, the presence is more frequent in the regions.

The econometric results offer important evidence regarding employment multipliers in the context of complexity. First, the hypothesis is confirmed that the multiplier of the complex sector is the largest. Second, the heterogeneity of multipliers is notable when considering micro-regions with different levels of complexity. Third, it appears that in less complex regions, the complex sector does not have a significant effect on itself and on the non-complex sector. The most positive effects in these regions are reserved for the influence that the non-complex sector has on itself. Fourth, the same cannot be said for regions that are already complex, as they are where the complex sector exerts the most prominent influence on the labor market, concentrating the largest multipliers for this sector. In short, the econometric results demonstrate that the *bad news* for less complex regions is a complex sector incapable of generating jobs, while the *good news* for complex regions is a complex sector capable of generating between 1.06 and 1.46 jobs in the same sector and between 1.71 and 3.25 in the non-complex sector of the economy.

These findings quantify one of the implications of the uneven development faced by Brazil's regions in terms of complexity. It shows that the development path followed by less complex regions limits their diversification opportunities to less complex sectors only. As a result, these regions are unable to develop a complex sector that can have positive spillover effects on the rest of the economy. This lack of diversification leads to almost non-existent multipliers for the complex sector in these regions. On the other hand, regions that are already complex benefit from related diversification, as their productive structure concentrates similar capabilities for the production of other complex activities. This dynamic results in a complex sector that can exert a significant influence on the overall economy. Hence, the multipliers of the complex sector in these regions are substantial.

While this work provides valuable contributions to the literature, it is important to acknowledge its limitations. The classification of sectors as complex or non-complex is a subjective decision and may influence the magnitude of the multipliers. Choosing a more inclusive classification would likely decrease the multipliers' magnitude and favor Medium-Low complexity regions, as evidenced by robustness tests in the annex. Moreover, the level of aggregation of economic activities can also impact the multipliers, with higher levels potentially yielding different magnitudes. Additionally, it is crucial to note that the estimates represent average impacts and do not account for variations in local conditions within each region group. Local factors and the specific characteristics of individual activities within the complex or non-complex sectors can also influence the multiplier effect. Therefore, the multipliers are still sensitive to local and sector-specific factors.

Conclusions

This master's thesis aimed to study the unequal nature of related diversification in Brazilian micro-regions. The prominent path dependence of the diversification process widens the gap between regions, as complex sector production is limited to already developed regions. This dynamic perpetuates a feedback loop of inequality (PINHEIRO et al., 2022), where relatedness serves not only as a driver of diversification but also as a significant contributor to regional disparities. As a result, it is argued that relatedness is *good news* for some and *bad news* for others.

The notion that Economic Geography should be regarded as an Evolutionary Science (BOSCHMA; FRENKEN, 2006) provides the theoretical groundwork for characterizing regional economic diversification. Chapter 1 builds on this foundation to theoretically demonstrate that related diversification also contributes to exacerbating inequality between regions. In this context, we contend that the literature on this aspect of regional development is scarce, leaving several unanswered questions. For instance, there is a lack of comprehensive research on the influence of sectoral complexity on the entry and exit probabilities of firms and how this influence varies across different levels of regional complexity. Additionally, little is known about the implications of diversification targeting more complex sectors and its impact on local economic drivers such as job creation. These gaps guided the discussions presented in Chapters 2 and 3.

Chapter 2 aimed to explain empirically why news is good for some and bad for others. In practical terms, we assessed the influence of relatedness and complexity on firm entry and exit, considering regions with different complexity levels. To capture the heterogeneity of this effect, we segmented regions based on their ECI values into four groups: Low, Medium-Low, Medium-High, and High. Employing the econometric framework proposed by Neffke, Henning and Boschma (2011), we used the occurrence of firm entry and exit from local portfolios, measured by the RCA value, as dependent variables. Using data from the formal labor market for three different years (2009, 2014, 2019), we confirmed our assumed hypotheses through three estimation strategies (OLS, logit, and probit). The results are as follows:

i) In more complex regions, the sector's complexity increases its probability of entering the portfolio and has little influence on its probability of exit.

ii) In less complex regions, the sector's complexity decreases its probability of entry and significantly increases its probability of exit.

These two hypothesis had not been thoroughly tested in the literature, representing the primary contributions of this chapter. These findings enrich the theoretical field, which considers the dynamics of firm entry and exit as a key explanation for regional diversification, with analyses conducted on Swedish (NEFFKE; HENNING; BOSCHMA, 2011), American (ESSLETZBICHLER, 2015; BOSCHMA; BALLAND; KOGLER, 2015) and Brazilian regions (FREITAS, 2019; FRANCOSO; BOSCHMA; VONORTAS, 2022). Additionally, this chapter extends this literature by focusing on sectoral and regional economic complexity, a relatively recent approach not extensively studied in previous research.

Additionally, these findings pave the way for essential research agendas. The tests conducted in Chapter 2 focus solely on Brazilian micro-regions, which possess distinct characteristics compared to European regions, for instance. Future studies should include samples from diverse regions to assess whether the identified dynamics hold true. Moreover, while the estimates reveal the uneven nature of related diversification, conducting case studies could provide more comprehensive insights into the challenges less complex regions encounter when diversifying into complex sectors.

Chapter 3 was aimed at assessing how good or how bad the news is for regions in terms of employment. This chapter presents a groundbreaking adaptation of Moretti and Thulin (2013)‘s methodology, originally used to measure employment multipliers for the tradable sector in Sweden, to analyze the diverse multipliers of both complex and non-complex sectors within micro-regions in Brazil. This adaptation enables the calculation of the number of jobs created in each sector for each additional job generated in either the complex or non-complex sector of the economy. To test the hypotheses, the classification of regions by complexity, as established in Chapter 2, was maintained.

The econometric tests followed the approach proposed by Moretti and Thulin (2013), using shift-share instruments. Moreover, an additional instrument was introduced to account for the structural changes that occurred during the period (STILWELL, 1969). Employing the same time frames as the previous chapter, the tests conducted in Chapter 3, utilizing both instruments, demonstrated the robustness of the results and confirmed the following assumptions:

i) The complex sector has higher employment multipliers.

ii) The multipliers are heterogeneous in relation to the complexity of the regions.

iii) Complex sector multipliers are higher in already complex regions.

The first hypothesis posits that the presence of complex and less common activities (manufacturing and financial services) in the local economy will stimulate demand for less complex and more common activities (basic services, food and beverage trade, sale of construction materials). The other two hypotheses link higher regional complexity to more established institutions, increased labor competition among sectors, and greater labor mobility, rendering the labor supply more responsive to changes in the complex sector.

These findings have important contributions in several senses. Firstly, this study enriches the literature on local employment multipliers by proposing an innovative adaptation of the conceptual framework, incorporating the concepts formulated by the economic complexity approach. Secondly, the chapter enhances the complexity literature by evaluating the implications of diversifying into more complex sectors on regional job creation. Thirdly, Chapter 3 introduces a general and easily replicable model for assessing a practical effect of complexity, which holds significant value in providing information to policymakers.

The contributions made in Chapter 3 also lay the foundation for significant research agendas. The adaptation of the conceptual framework to calculate complex and non-complex sector multipliers represents an innovative approach that can be further refined to ensure the robustness of the results. For instance, future studies could explore alternative criteria for categorizing sectors, moving beyond the current distribution-based approach. As the methodology hinges on the duality of sectors, further theoretical advancements are necessary to precisely define what constitutes complex and non-complex sectors. Moreover, the results reveal that the positive and significant effects of the complex sector are observed only in regions that are already complex, raising the question of when this effect begins to manifest as complexity increases. This aspect merits further investigation in forthcoming research.

The findings of this master's thesis hold significant policy implications. As highlighted in Chapter 2, policymakers must address the trend of regional development disparities carefully. Specifically, for less complex regions, policies should focus on creating opportunities to enhance their capabilities for diversifying into complex sectors that are not too distant from their existing portfolios. This approach can break the cycle of retrograde development faced by less complex regions. Moreover, Chapter 3's results reveal that a complex sector's effectiveness in generating jobs is contingent on a minimum degree of regional complexity. Therefore, policies aimed at increasing a region's complexity can trigger the multiplier effect of the complex sector, mitigating the challenges posed by divergent development. Consequently, promoting unrelated diversification through targeted policies presents an alternative approach to redirect the diversification trajectory of less complex regions.

Finally, it is important to address some general limitations of this master's thesis that should be taken into account in future research. Firstly, the analyses are solely based on the formal labor market data available in the RAIS database. As the Brazilian informal labor market has experienced growth in recent years, it is essential to include it in the analysis, despite the challenges of obtaining information on this sector. Secondly, the estimations performed in this study do not account for spatial aspects. The influence of neighboring regions on the diversification process is not considered, even though extra-regional connections can be significant determinants of regional diversification. Despite these limitations, this master's thesis provides valuable contributions to the literature by exploring a theme that is often overlooked, shedding light on important research areas for future studies.

Bibliography

BALASSA, B. Trade Liberalization and Revealed Comparative Advantage. The Manchester School of Economic and Social Studies, v. 33, p. 99–123, 1965.

BALLAND, P.-A.; BOSCHMA, R. Complementary interregional linkages and Smart Specialisation: an empirical study on European regions. *Regional Studies*, v. 55, n. 6, p. 1059–1070, 2021.

BALLAND, P.-A. et al. Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, p. 1–17, 2018.

BALLAND, P.-A. et al. Complex economic activities concentrate in large cities. *Nature Human Behaviour*, v. 4, p. 248–254, 2020.

BALLAND, P.-A.; RIGBY, D. The Geography of Complex Knowledge. *Economic Geography*, v. 93, n. 1, p. 1–23, 2017.

BARTIK, T. J. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo: W.E. Upjohn Institute for Employment Research, 1991.

BLASIO, G. de; MENON, C. Local Effects of Manufacturing Employment Growth in Italy. *Giornale degli Economisti*, v. 70, n. 3, p. 101–112, 2011.

BOSCHMA, R. Proximity and Innovation: A Critical Assessment. *Regional Studies*, v. 39, n. 1, p. 61–74, 2005.

BOSCHMA, R. Relatedness as driver of regional diversification: a research agenda. *Regional Studies*, v. 51, n. 3, p. 351–364, 2017.

BOSCHMA, R. Designing Smart Specialization Policy: relatedness, unrelatedness, or what? Papers in Evolutionary Economic Geography (PEEG), n. 2128, 2021.

BOSCHMA, R.; BALLAND, P.-A.; KOGLER, D. F. 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*, v. 24, n. 1, p. 223–250, 2015.

BOSCHMA, R.; FRENKEN, K. Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography*, v. 6, n. 3, p. 273–302, 2006.

BOSCHMA, R.; FRENKEN, K. The emerging empirics of evolutionary economic geography. *Journal of Economic Geography*, v. 11, n. 2, p. 295–307, 2011.

BOSCHMA, R.; HEIMERIKS, G.; BALLAND, P.-A. Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy*, v. 43, n. 1, p. 107–114, 2014.

BOSCHMA, R.; IAMMARINO, S. Related Variety, Trade Linkages, and Regional Growth in Italy. *Economic Geography*, v. 85, n. 3, p. 289–311, 2009.

BOSCHMA, R.; LAMBOOY, J. Evolutionary economics and economic geography. *Journal of Evolutionary Economics*, v. 9, n. 4, p. 411–429, 1999.

BOSCHMA, R.; MINONDO, A.; NAVARRO, M. The Emergence of New Industries at the Regional Level in Spain: A Proximity Approach Based on Product Relatedness. *Economic Geography*, v. 89, n. 1, p. 29–51, 2013.

CHU, L. K.; HOANG, D. P. How does economic complexity influence income inequality? New evidence from international data. *Economic Analysis and Policy*, v. 68, n. C, p. 44–57, 2020.

CIMINI, F. et al. A armadilha da baixa complexidade em Minas Gerais: o desafio da sofisticação econômica em um estado exportador de commodities. *Revista Brasileira de Inovação*, v. 17, p. 33–62, 2017.

DIJK, J. J. V. Local multipliers, mobility, and agglomeration economies. Discussion Paper No. 771, University of Oxford, 2015.

DIJK, J. J. V. Local employment multipliers in U.S. cities. *Journal of Economic Geography*, v. 17, n. 2, p. 465–487, 2017.

EFRON, B.; TIBSHIRANI, R. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

ESSLETZBICHLER, J. Relatedness, Industrial Branching and Technological Cohesion in US Metropolitan Areas. *Regional Studies*, v. 49, n. 5, p. 752–766, 2015.

FAGGIO, G.; OVERMAN, H. The effect of public sector employment on local labour markets. *Journal of Urban Economics*, v. 79, n. C, p. 91–107, 2014.

FLORENCE, S. P. *Investment, location and size of plant*. Cambridge: Cambridge University Press, 1948.

FRANCOSO, M. S.; BOSCHMA, R.; VONORTAS, N. Regional diversification in Brazil: the role of relatedness and complexity. Papers in Evolutionary Economic Geography (PEEG), 2022.

FREITAS, E. *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, 2019.

FREITAS, E.; PAIVA, E. Diversificação e sofisticação das exportações: uma aplicação do Product Space aos dados do Brasil. *Revista Econômica do Nordeste*, v. 46, p. 79–98, 2015.

FRENKEN, K.; OORT, F. V.; VERBURG, T. Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies*, v. 41, n. 5, p. 685–697, 2007.

FURTADO, C. *Development and Underdevelopment*. Berkeley: University of California Press, 1964.

GEROLIMETTO, M.; MAGRINI, S. Spatial analysis of employment multilpliers in Spanish labor markets. *RIEDS - Rivista Italiana di Economia, Demografia e Statistica - The Italian Journal of Economic, Demographic and Statistical Studies*, v. 68, n. 3-4, p. 87–94, 2014.

GLAESER, E. et al. Growth in Cities. *Journal of Political Economy*, v. 100, n. 6, p. 1126–1152, 1992.

GOOS, M.; KONINGS, J.; VANDEWEYER, M. Local high-tech job multipliers in Europe. *Industrial and Corporate Change*, v. 27, n. 4, p. 639–655, 2018.

GUTIERREZ-POSADA, D. et al. Creative Clusters and Creative Multipliers: Evidence from UK Cities. *Economic Geography*, v. 99, n. 1, p. 1–24, 2023.

HARTMANN, D. et al. International trade, development traps, and the core-periphery structure of income inequality. *EconomiA*, v. 21, p. 1441–1452, 2020.

HARTMANN, D.; BEZERRA, M.; PINHEIRO, F. L. Identifying Smart Strategies for Economic Diversification and Inclusive Growth in Developing Economies. The Case of Paraguay. SSRN, 2019.

HARTMANN, D. et al. Linking Economic Complexity, Institutions, and Income Inequality. *World Development*, v. 93, p. 75–93, 2017.

HARTMANN, D. et al. The structural constraints of income inequality in Latin America. arXiv.org, n. 1701.03770, 2017.

HARTMANN, D.; PINHEIRO, F. Economic complexity and inequality at the national and regional level. arXiv.org, n. 2206.00818, 2022.

HAUSMANN, R.; CHAUVIN, J. Moving to the Adjacent Possible: Discovering Paths for Export Diversification in Rwanda. Center for International Development (CID) Faculty Working Paper, n. 294, 2015.

HAUSMANN, R. et al. *The Atlas of Economic Complexity: Mapping Paths to Prosperity.* 1. ed. Cambridge: The MIT Press, 2014. v. 1.

HAUSMANN, R.; KLINGER, B. The Structure of the Product Space and the Evolution of Comparative Advantage. CID Working Papers, n. 146, 2007.

HAUSMANN, R.; SANTOS, M. A.; OBACH, J. Appraising the Economic Potential of Panama Policy Recommendations for Sustainable and Inclusive Growth. CID Working Papers, Harvard University, n. 334, 2017.

HENDERSON, J. V.; KUNCORO, A.; TURNER, M. Industrial Development in Cities. *Journal of Political Economy*, v. 103, n. 5, p. 1067–1090, 1995.

HERNANDEZ, N.; ROJAS, I. Employment local multipliers in mexico. Sobre México. Revista de Economía, v. 1, p. 5–37, 2020.

HIDALGO, C. et al. The Principle of Relatedness: Proceedings of the Ninth International Conference on Complex Systems. p. 451–457, 2018.

HIDALGO, C.; HAUSMANN, R. The Building Blocks of Economic Complexity. *Proceedings of the National Academy of Sciences of the United States of America*, v. 106, p. 10570–5, 2009.

HIDALGO, C. et al. The Product Space Conditions the Development of Nations. *Science*, v. 317, p. 482–487, 2007.

HIRSCHMAN, A. *The Strategy of Economic Development.* New Haven: Yale University Press, 1958.

IAMMARINO, S.; RODRIGUEZ-POSE, A.; STORPER, M. Regional inequality in Europe: Evidence, theory and policy implications. *Journal of Economic Geography*, v. 19, p. 273–298, 2019.

IBGE. Divisão regional do brasil em mesorregiões e microrregiões geográficas. v. 1, 1990.

JACOBS, J. *The Economy of Cities*. New York: Vintage, 1969.

KAZEKAMI, S. Local Multipliers, Mobility, and Agglomeration Economies. *Industrial Relations: A Journal of Economy and Society*, v. 56, n. 3, p. 489–513, 2017.

KOGLER, D.; RIGBY, D.; TUCKER, I. Mapping Knowledge Space and Technological Relatedness in US Cities. *European Planning Studies*, v. 21, n. 9, p. 1374–1391, 2013.

KRUGMAN, P. Increasing Returns and Economic Geography. Journal of Political Economy, n. 3, p. 483–499, 1991.

LAPATINAS, A. Economic complexity and human development: a note. Economics Bulletin, v. 36, n. 3, p. 1441–1452, 2016.

LEE, K.; VU, T. Economic complexity, human capital and income inequality: a cross-country analysis. *The Japanese Economic Review*, v. 71, n. 4, p. 695–718, 2020.

LEE, N. Are Innovative Regions More Unequal? Evidence from Europe. *Environment and Planning C: Government and Policy*, v. 29, n. 1, p. 2–23, 2011.

LEE, N.; CLARKE, S. Do low-skilled workers gain from high-tech employment growth? High-technology multipliers, employment and wages in Britain. *Research Policy*, v. 48, n. 9, 2019.

LEE, N.; RODRIGUEZ-POSE, A. Innovation and spatial inequality in Europe and USA. *Journal of Economic Geography*, v. 13, n. 1, p. 1–22, 2013.

LEWIS, A. *The Theory of Economic Growth*. London: Allen & Unwin, 1955.

LOYO, A. O. L.; MOISÉS, A. R. F.; MENDES, V. Impacto de mudanças no emprego no setor público sobre o mercado de trabalho local: evidências para as mesorregiões brasileiras de 2003 a 2010. *Estudos Econômicos*, v. 48, p. 77–106, 2018.

MACEDO, G.; MONASTERIO, L. Local multiplier of industrial employment: Brazilian mesoregions (2000-2010). *Brazilian Journal of Political Economy*, v. 36, n. 4, p. 827–839, 2016.

MARCO, R.; LLANO, C.; PÉREZ-BALSALOBRE, S. Economic complexity, environmental quality and income equality: A new trilemma for regions? *Applied Geography*, p. 102646, 2022.

MARSHALL, A. *Principles of economics: An introductory volume*. London: Macmillan, 1920.

MORAIS, B. M.; SWART, J.; JORDAAN, J. A. Economic Complexity and Inequality: Does Regional Productive Structure Affect Income Inequality in Brazilian States? *Sustainability*, v. 13, n. 2, p. 1–23, 2021.

MORETTI, E. Local Multipliers. *American Economic Review*, v. 100, n. 2, p. 373–377, 2010.

MORETTI, E.; THULIN, P. Local multipliers and human capital in the United States and Sweden. *Industrial and Corporate Change*, v. 22, n. 1, p. 339–362, 2013.

MYRDAL, G. *Economic Theory and Underdeveloped Regions*. London: G. Duckworth, 1957.

NEFFKE, F.; HENNING, M.; BOSCHMA, R. How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography*, v. 87, n. 3, p. 237–265, 2011.

NELSON, R. R.; WINTER, S. G. *An Evolutionary Theory of Economic Change*. Cambridge, MA and London: The Belknap Press, 1982.

PINHEIRO, F. et al. Shooting Low or High: Do Countries Benefit from Entering Unrelated Activities? Papers in Evolutionary Economic Geography (PEEG), n. 1807, p. 1–44, 2018.

PINHEIRO, F. et al. The Dark Side of the Geography of Innovation. Relatedness, Complexity, and Regional Inequality in Europe. Papers in Evolutionary Economic Geography (PEEG), n. 2202, 2022.

PREBISCH, R. *The Economic Development of Latin America and Its Principal Problems*. New York: United Nations Department of Economic Affairs, Economic Commission for Latin America (ECLA), 1950.

QUEIROZ, A. R.; ROMERO, J. P.; FREITAS, E. E. Economic complexity and employment in Brazilian states. *CEPAL Review*, p. 177–196, 2023.

RIGBY, D. L. Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes. *Regional Studies*, v. 49, n. 11, p. 1922–1937, 2015.

ROCHA, R. M.; ARAÚJO, B. C. Local multiplier effect of the tradable sector on the Brazilian labor market. *Letters in Spatial and Resource Sciences*, v. 14, n. 3, p. 269–286, 2021.

ROMERO, J. et al. Economic complexity and regional development: evidence from brazilian municipalities. *Proceedings of the EAEPE Conference*, p. 1–25, 2022.

ROMERO, J.; SILVEIRA, F. Mudança estrutural e complexidade econômica: identificando setores promissores para o desenvolvimento dos estados brasileiros. In: LEITE, M. V. C. (Org.). *Alternativas para o desenvolvimento brasileiro: novos horizontes para a mudança estrutural com igualdade*. 201. ed. Santiago: Nações Unidas, Comissão Econômica para a América Latina e o Caribe (CEPAL), 2019. p. 137–160.

SBARDELLA, A.; PUGLIESE, E.; PIETRONERO, L. Economic development and wage inequality: A complex system analysis. *PLoS ONE*, v. 12, n. 9, p. 1–26, 2017.

SCHUMPETER, J. *The Theory of Economic Development*. Cambridge, MA: Harvard University Press, 1934.

SEPEHRDOUST, H.; TARTAR, M.; GHOLIZADEH, A. Economic complexity, scientific productivity and income inequality in developing economies. *Economics of Transition and Institutional Change*, v. 30, n. 4, p. 737–752, 2022.

SHANNON, C. E. A mathematical theory of communication. *Bell System Technical Journal*, v. 27, p. 379–423, 1948.

SIMOES, R.; FREITAS, E. Urban Attributes and Regional Differences in Productivity: Evidence from the External Economics of Brazilian Micro-regions from 2000 - 2010. Journal of Economic and Financial Studies (JEFS), v. 2, n. 1, p. 27–39, 2014.

STILWELL, F. J. B. Regional Growth and Structural Adaptation. Urban Studies, v. 6, n. 2, p. 162–178, 1969.

VEBLEN, T. Why is Economics Not an Evolutionary Science? *The Quarterly Journal of Economics*, v. 12, n. 4, p. 373–397, 1898.

WANG, T.; CHANDA, A. Manufacturing growth and local employment multipliers in China. *Journal of Comparative Economics*, v. 46, n. 2, p. 515–543, 2018.

WOOLDRIDGE, J. M. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA and London: The MIT Press, 2010.

Annex

ANNEX A – Chapter 2

Table 12 – Emergence of new activities - OLS models

	Dependent variable:				
			Entry		
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	0.646***	1.373***	0.878***	0.858***	1.048***
	(0.029)	(0.130)	(0.052)	(0.060)	(0.085)
PCI	−0.049**	−0.020	−0.046**	0.016	0.062***
	(0.019)	(0.016)	(0.022)	(0.024)	(0.024)
Diversity	−0.005***	−0.016***	−0.014***	−0.032***	−0.031*
	(0.001)	(0.003)	(0.002)	(0.005)	(0.019)
Log(GDPpc)	−0.002	0.001	0.001	0.004	−0.032*
	(0.002)	(0.004)	(0.002)	(0.005)	(0.017)
Log(Population)	−0.010***	−0.001	−0.005***	−0.009***	−0.050***
	(0.002)	(0.002)	(0.001)	(0.003)	(0.009)
Region Productivity	−0.004	−0.015***	−0.006	0.005	0.037*
	(0.003)	(0.005)	(0.004)	(0.008)	(0.022)
Region HC	−0.026***	−0.014*	−0.039***	−0.060***	0.058
	(0.008)	(0.008)	(0.008)	(0.023)	(0.074)
Incentives	0.003	−0.002	0.005**	0.006	0.012
	(0.002)	(0.003)	(0.002)	(0.007)	(0.030)
Credit	−0.000**	0.000	−0.000**	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sector Size	−0.004***	−0.001	−0.004**	−0.007***	−0.010***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
CL	−0.084***	−0.018**	−0.071***	−0.134***	−0.180***
	(0.009)	(0.007)	(0.010)	(0.013)	(0.019)
Sector HC	0.008	0.012	−0.002	0.022	0.036
	(0.011)	(0.011)	(0.013)	(0.014)	(0.026)
Sector Productivity	−0.002	0.003	−0.0001	−0.005	−0.002
	(0.003)	(0.003)	(0.003)	(0.004)	(0.008)
Constant	0.254***	0.109***	0.196***	0.316***	0.606***
	(0.031)	(0.035)	(0.027)	(0.059)	(0.154)
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
R^2	0.049	0.046	0.051	0.046	0.062
Residual Std. Error	0.212	0.137	0.199	0.250	0.284
F Statistic	266.430***	29.185***	169.961***	60.857***	17.914***

Note: *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.

Table 13 – Emergence of new activities - Probit models

	Dependent variable:				
	Entry				
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	4.931***	12.142***	6.250***	6.918***	8.229***
	(0.285)	(1.488)	(0.503)	(0.617)	(0.757)
PCI	−1.045***	−2.447***	−1.496***	−0.090	0.361**
	(0.205)	(0.345)	(0.235)	(0.234)	(0.168)
Diversity	0.020	−0.067	−0.050**	−0.269***	−0.301**
	(0.014)	(0.049)	(0.020)	(0.044)	(0.137)
Log(GDPpc)	−0.024	−0.042	0.010	0.035	−0.229*
	(0.024)	(0.093)	(0.022)	(0.042)	(0.125)
Log(Population)	−0.097***	0.070**	−0.017	−0.074***	−0.409***
	(0.018)	(0.027)	(0.013)	(0.025)	(0.078)
Region Productivity	−0.055	−0.326***	−0.106**	0.036	0.220
	(0.041)	(0.119)	(0.042)	(0.074)	(0.177)
Region HC	−0.251**	−0.210	−0.286***	−0.524***	0.477
	(0.099)	(0.139)	(0.094)	(0.202)	(0.569)
Incentives	0.033	−0.102	0.049**	0.061	0.094
	(0.024)	(0.062)	(0.024)	(0.056)	(0.228)
Credit	−0.000*	0.00000	−0.000*	−0.000	−0.000
	(0.000)	(0.00000)	(0.000)	(0.000)	(0.000)
Sector Size	−0.076***	−0.070***	−0.082***	−0.076***	−0.082***
	(0.013)	(0.027)	(0.015)	(0.015)	(0.022)
CL	−0.982***	−0.687***	−0.985***	−1.169***	−1.303***
	(0.101)	(0.174)	(0.111)	(0.109)	(0.130)
Sector HC	0.037	0.093	−0.133	0.278**	0.387**
	(0.114)	(0.212)	(0.139)	(0.123)	(0.185)
Sector Productivity	−0.071**	−0.043	−0.061	−0.071*	0.004
	(0.035)	(0.067)	(0.039)	(0.037)	(0.064)
Constant	0.893**	0.564	0.547*	0.903*	3.383***
	(0.347)	(0.738)	(0.326)	(0.521)	(1.225)
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 R^2	0.12	0.2	0.14	0.1	0.11
Log Likelihood	−112,270.500	−5,637.414	−60,553.320	−35,389.430	−8,407.127
Akaike Inf. Crit.	224,793.000	11,510.830	121,356.600	71,024.870	17,036.250

Note: *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.

Table 14 – Exit of activities - OLS models

	Dependent variable:				
			Entry		
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	−1.334***	−1.931**	−2.094***	−1.705***	−1.592***
	(0.082)	(0.893)	(0.187)	(0.225)	(0.203)
PCI	0.368***	1.437***	0.557***	0.123**	−0.058
	(0.043)	(0.231)	(0.067)	(0.051)	(0.052)
Diversity	0.027***	0.032	0.059***	0.118***	0.049*
	(0.006)	(0.033)	(0.009)	(0.017)	(0.025)
Log(GDPpc)	0.005	−0.110***	−0.002	0.001	0.049**
	(0.008)	(0.037)	(0.009)	(0.015)	(0.019)
Log(Population)	0.023***	−0.011	0.002	0.009	0.078***
	(0.005)	(0.019)	(0.006)	(0.009)	(0.017)
Region Productivity	−0.021	0.215***	−0.009	−0.049	−0.134***
	(0.016)	(0.067)	(0.019)	(0.030)	(0.034)
Region HC	0.103***	−0.084	0.130***	0.174**	0.270*
	(0.040)	(0.115)	(0.046)	(0.084)	(0.145)
Incentives	−0.009	−0.005	−0.023*	0.009	−0.030
	(0.011)	(0.043)	(0.013)	(0.024)	(0.054)
Credit	0.000**	−0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sector Size	−0.054***	−0.056**	−0.055***	−0.053***	−0.042***
	(0.006)	(0.023)	(0.007)	(0.007)	(0.007)
CL	0.290***	0.236*	0.268***	0.365***	0.548***
	(0.039)	(0.132)	(0.047)	(0.042)	(0.049)
Sector HC	0.112**	−0.013	0.129*	0.069	0.134**
	(0.053)	(0.179)	(0.069)	(0.054)	(0.053)
Sector Productivity	0.023*	0.074	0.025	0.018	0.003
	(0.013)	(0.049)	(0.018)	(0.015)	(0.017)
Constant	−0.185**	−1.501***	−0.079	−0.232	−0.322
	(0.093)	(0.501)	(0.126)	(0.181)	(0.321)
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
R^2	0.082	0.183	0.116	0.068	0.066
Residual Std. Error	0.429	0.435	0.429	0.429	0.400
F Statistic	71.541***	6.112***	49.956***	21.891***	8.182***

Note: *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.

Table 15 – Exit of activities - Probit models

	Dependent variable:				
			Entry		
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	−4.507***	−7.094**	−7.186***	−5.542***	−6.272***
	(0.291)	(2.988)	(0.608)	(0.754)	(0.747)
PCI	1.102***	4.445***	1.626***	0.358**	−0.251
	(0.140)	(0.820)	(0.220)	(0.161)	(0.179)
Diversity	0.102***	0.125	0.210***	0.377***	0.228**
	(0.018)	(0.108)	(0.028)	(0.055)	(0.097)
Log(GDPpc)	0.012	−0.365***	−0.005	0.002	0.159**
	(0.025)	(0.119)	(0.030)	(0.047)	(0.068)
Log(Population)	0.081***	−0.035	0.013	0.037	0.304***
	(0.017)	(0.063)	(0.019)	(0.030)	(0.064)
Region Productivity	−0.066	0.692***	−0.043	−0.152	−0.483***
	(0.051)	(0.220)	(0.062)	(0.096)	(0.125)
Region HC	0.338***	−0.219	0.446***	0.547**	0.877*
	(0.125)	(0.383)	(0.148)	(0.272)	(0.512)
Incentives	−0.031	−0.011	−0.080*	0.030	−0.157
	(0.034)	(0.140)	(0.044)	(0.077)	(0.192)
Credit	0.000***	−0.000	0.000	0.000	0.000**
	(0.000)	(0.00000)	(0.000)	(0.000)	(0.000)
Sector Size	−0.174***	−0.176**	−0.177***	−0.166***	−0.151***
	(0.019)	(0.075)	(0.024)	(0.021)	(0.025)
CL	0.932***	0.863*	0.853***	1.149***	1.975***
	(0.123)	(0.445)	(0.150)	(0.131)	(0.169)
Sector HC	0.361**	−0.111	0.379*	0.249	0.498***
	(0.166)	(0.582)	(0.222)	(0.169)	(0.191)
Sector Productivity	0.074*	0.244	0.087	0.054	0.009
	(0.041)	(0.155)	(0.056)	(0.047)	(0.059)
Constant	−2.211***	−6.441***	−1.773***	−2.325***	−3.064***
	(0.298)	(1.627)	(0.399)	(0.583)	(1.140)
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 R^2	0.07	0.15	0.1	0.06	0.07
Log Likelihood	−54,754.510	−1,655.946	−25,881.640	−20,164.810	−6,223.357
Akaike Inf. Crit.	109,761.000	3,529.891	52,013.280	40,575.630	12,668.710

Note: *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.

Table 16 – Emergence of new activities - ECI Groups by Quartile

	Logit - Dependent variable:				
			Entry		
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	9.864***	16.416***	14.491***	15.308***	13.459***
	(0.629)	(1.914)	(1.473)	(1.591)	(0.796)
PCI	−2.232***	−5.750***	−3.639***	−1.375**	0.276
	(0.426)	(0.643)	(0.549)	(0.593)	(0.335)
Diversity	0.054*	0.033	−0.113**	−0.317***	−0.437***
	(0.031)	(0.067)	(0.055)	(0.087)	(0.102)
Log(GDPpc)	−0.069	−0.120	−0.061	0.207***	−0.044
	(0.053)	(0.114)	(0.052)	(0.073)	(0.095)
Log(Population)	−0.217***	0.002	−0.009	−0.086*	−0.433***
	(0.039)	(0.052)	(0.037)	(0.047)	(0.058)
Region Productivity	−0.103	−0.270	−0.188	−0.148	0.120
	(0.091)	(0.187)	(0.139)	(0.138)	(0.146)
Region HC	−0.497**	−0.484*	−0.517*	−1.268***	−0.373
	(0.221)	(0.282)	(0.287)	(0.393)	(0.480)
Incentives	0.061	−0.090	0.021	0.105	−0.019
	(0.053)	(0.099)	(0.076)	(0.101)	(0.130)
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***	−0.191***	−0.169***	−0.199***	−0.160***
	(0.029)	(0.051)	(0.043)	(0.035)	(0.029)
CL	−2.047***	−2.107***	−2.068***	−2.192***	−2.418***
	(0.224)	(0.334)	(0.280)	(0.251)	(0.205)
Sector HC	0.057	−0.384	−0.356	−0.102	0.588***
	(0.245)	(0.392)	(0.346)	(0.295)	(0.217)
Sector Productivity	−0.142*	−0.167	−0.160	−0.146	−0.100
	(0.076)	(0.124)	(0.101)	(0.090)	(0.069)
Constant	2.454***	2.265*	1.489	1.883*	4.858***
	(0.752)	(1.228)	(0.970)	(1.043)	(1.048)
Fixed effects - UF	Yes	Yes	Yes	Yes	Yes
Fixed effects - Activities	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.12	0.17	0.15	0.12	0.09
Observations	647,801	176,297	167,992	158,462	145,050
Log Likelihood	−112,609.600	−17,988.080	−25,462.940	−30,355.160	−36,129.490
Akaike Inf. Crit.	225,471.200	36,224.150	51,171.880	60,960.320	72,506.980

Note: *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.

Table 17 – Exit of activities - ECI Groups by Quartile

	Logit - Dependent variable:				
			Exit		
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	−7.752***	−9.675***	−13.844***	−16.845***	−8.943***
	(0.502)	(3.087)	(1.679)	(1.304)	(0.707)
PCI	1.847***	6.235***	3.301***	0.257	0.215
	(0.238)	(0.895)	(0.505)	(0.352)	(0.219)
Diversity	0.178***	0.216***	0.356***	0.681***	0.512***
	(0.031)	(0.079)	(0.075)	(0.082)	(0.081)
Log(GDPpc)	0.015	−0.268**	0.009	−0.075	0.037
	(0.043)	(0.135)	(0.067)	(0.075)	(0.069)
Log(Population)	0.132***	−0.131*	−0.067	0.117**	0.241***
	(0.028)	(0.076)	(0.045)	(0.046)	(0.052)
Region Productivity	−0.115	0.479**	−0.165	−0.221	−0.370***
	(0.084)	(0.195)	(0.161)	(0.141)	(0.131)
Region HC	0.580***	0.054	0.802*	1.044**	1.290***
	(0.209)	(0.434)	(0.418)	(0.408)	(0.477)
Incentives	−0.051	−0.127	−0.117	0.010	0.018
	(0.056)	(0.183)	(0.101)	(0.115)	(0.126)
Credit	0.000***	0.000	0.000*	−0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sector Size	−0.298***	−0.265***	−0.341***	−0.290***	−0.278***
	(0.033)	(0.077)	(0.053)	(0.036)	(0.033)
CL	1.560***	1.264**	1.408***	1.807***	2.352***
	(0.206)	(0.521)	(0.286)	(0.235)	(0.209)
Sector HC	0.595**	0.596	0.733	0.618*	0.472*
	(0.278)	(0.654)	(0.477)	(0.322)	(0.252)
Sector Productivity	0.132*	0.195	0.238**	0.097	0.081
	(0.069)	(0.184)	(0.119)	(0.095)	(0.068)
Constant	−3.642***	−4.786***	−1.880*	−3.350***	−5.162***
	(0.501)	(1.388)	(0.982)	(0.968)	(0.801)
Fixed effects - UF	Yes	Yes	Yes	Yes	Yes
Fixed effects - Activities	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.07	0.13	0.12	0.08	0.06
Observations	99,919	10,633	18,938	28,468	41,880
Log Likelihood	−54,737.680	−5,903.602	−10,106.110	−15,494.630	−22,307.470
Akaike Inf. Crit.	109,727.400	12,055.200	20,458.230	31,239.250	44,862.930

Note: *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.

Table 18 – Emergence of new activities - Relatedness measured by Co-occupation

	Logit - Dependent variable:				
			Entry		
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	6.659***	12.482***	5.972***	8.580***	11.775***
	(0.442)	(1.924)	(0.725)	(0.584)	(0.977)
PCI	−3.312***	−8.650***	−5.255***	−1.797***	1.051***
	(0.404)	(0.731)	(0.422)	(0.369)	(0.310)
Diversity	0.128***	0.188**	0.114***	−0.248***	−0.458**
	(0.027)	(0.084)	(0.035)	(0.062)	(0.222)
Log(GDPpc)	−0.025	−0.214	−0.006	0.072	−0.184
	(0.045)	(0.217)	(0.044)	(0.065)	(0.195)
Log(Population)	−0.092***	0.155**	0.040	−0.043	−0.543***
	(0.030)	(0.062)	(0.026)	(0.041)	(0.116)
Region Productivity	−0.224***	−0.887***	−0.390***	−0.126	0.050
	(0.085)	(0.284)	(0.097)	(0.128)	(0.275)
Region HC	−0.374**	−0.315	−0.266	−0.785**	1.577*
	(0.184)	(0.289)	(0.195)	(0.329)	(0.909)
Incentives	0.055	−0.236	0.056	0.145	0.116
	(0.046)	(0.145)	(0.049)	(0.093)	(0.333)
Credit	−0.000	0.000	−0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sector Size	−0.165***	−0.191***	−0.192***	−0.147***	−0.148***
	(0.031)	(0.066)	(0.036)	(0.036)	(0.043)
CL	−1.906***	−1.498***	−1.935***	−2.166***	−2.388***
	(0.220)	(0.419)	(0.253)	(0.219)	(0.257)
Sector HC	0.076	0.318	−0.300	0.598**	0.646*
	(0.255)	(0.510)	(0.315)	(0.268)	(0.330)
Sector Productivity	−0.112	−0.027	−0.095	−0.092	−0.008
	(0.075)	(0.172)	(0.088)	(0.077)	(0.106)
Constant	1.486**	2.946	1.678**	1.154	5.619***
	(0.662)	(1.803)	(0.714)	(0.842)	(1.804)
Fixed effects - UF	Yes	Yes	Yes	Yes	Yes
Fixed effects - Activities	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.12	0.2	0.14	0.09	0.1
Observations	647,801	71,305	392,864	153,505	30,127
Log Likelihood	−113,008.500	−5,619.874	−60,741.900	−35,631.590	−8,485.956
Akaike Inf. Crit.	226,268.900	11,475.750	121,733.800	71,509.170	17,193.910

Note: *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.

Table 19 – Exit of activities - Relatedness measured by Co-occupation

	Logit - Dependent variable:				
	Exit				
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	−5.672***	−8.493***	−6.539***	−5.576***	−8.210***
	(0.414)	(2.345)	(0.789)	(0.624)	(0.789)
PCI	2.438***	9.045***	4.501***	1.491***	−1.199***
	(0.266)	(1.235)	(0.338)	(0.274)	(0.315)
Diversity	0.114***	0.104	0.110***	0.354***	0.171
	(0.027)	(0.113)	(0.041)	(0.063)	(0.166)
Log(GDPpc)	−0.024	−0.576***	−0.007	−0.0002	0.153
	(0.041)	(0.211)	(0.046)	(0.072)	(0.121)
Log(Population)	−0.001	−0.083	−0.063**	−0.042	0.328***
	(0.025)	(0.097)	(0.029)	(0.040)	(0.081)
Region Productivity	−0.010	1.311***	0.122	−0.099	−0.685***
	(0.083)	(0.378)	(0.104)	(0.148)	(0.239)
Region HC	0.489***	−0.268	0.490**	0.805**	1.484
	(0.187)	(0.635)	(0.230)	(0.376)	(0.920)
Incentives	−0.068	0.018	−0.105	0.007	−0.313
	(0.054)	(0.238)	(0.066)	(0.108)	(0.288)
Credit	0.000	−0.000	0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sector Size	−0.317***	−0.317***	−0.314***	−0.302***	−0.281***
	(0.033)	(0.122)	(0.041)	(0.037)	(0.043)
CL	1.095***	1.220*	0.960***	1.364***	2.319***
	(0.216)	(0.739)	(0.269)	(0.224)	(0.275)
Sector HC	0.643**	−0.394	0.651*	0.475	1.007***
	(0.291)	(0.964)	(0.390)	(0.294)	(0.332)
Sector Productivity	0.086	0.299	0.078	0.040	0.015
	(0.070)	(0.252)	(0.091)	(0.082)	(0.096)
Constant	−1.930***	−10.496***	−2.146***	−2.034**	−2.068
	(0.478)	(2.712)	(0.663)	(0.938)	(1.617)
Fixed effects - UF	Yes	Yes	Yes	Yes	Yes
Fixed effects - Activities	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.07	0.16	0.1	0.05	0.06
Observations	99,919	3,065	47,326	36,775	12,753
Log Likelihood	−54,938.530	−1,651.807	−25,955.730	−20,243.360	−6,254.198
Akaike Inf. Crit.	110,129.100	3,521.614	52,161.450	40,732.720	12,730.400

Note: *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.

Table 20 – Emergence of new activities (2007-2017)

	Logit - Dependent variable:				
			Entry		
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	10.140***	22.029***	11.871***	16.284***	17.124***
	(0.656)	(6.102)	(1.146)	(1.434)	(1.898)
PCI	−1.761***	−6.720***	−3.055***	0.583*	1.589***
	(0.241)	(0.999)	(0.334)	(0.307)	(0.354)
Diversity	−0.004	0.369**	−0.091*	−0.737***	0.511
	(0.034)	(0.156)	(0.048)	(0.117)	(0.748)
Log(GDPpc)	0.119	−0.167	0.091	0.249	−0.289
	(0.081)	(0.419)	(0.069)	(0.167)	(0.532)
Log(Population)	−0.144***	0.029	0.013	−0.063	−1.105***
	(0.047)	(0.152)	(0.036)	(0.075)	(0.229)
Region Productivity	−0.250	−0.292	0.060	−0.462	−0.215
	(0.155)	(0.713)	(0.137)	(0.336)	(0.920)
Region HC	−0.940**	1.085	−0.384	−3.495***	7.382**
	(0.462)	(1.220)	(0.432)	(1.241)	(3.300)
Incentives	0.070	0.522	0.122	0.153	0.220
	(0.076)	(0.324)	(0.077)	(0.166)	(0.719)
Credit	−0.000	0.000	−0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sector Size	−0.189***	−0.277***	−0.195***	−0.173***	−0.220***
	(0.036)	(0.107)	(0.037)	(0.054)	(0.064)
CL	−1.795***	−1.197***	−1.949***	−2.016***	−1.813***
	(0.183)	(0.464)	(0.211)	(0.215)	(0.319)
Sector HC	−0.380	−1.313	−0.436	−0.436	−0.105
	(0.371)	(0.913)	(0.459)	(0.497)	(0.623)
Sector Productivity	−0.014	0.461	−0.048	0.039	−0.011
	(0.100)	(0.296)	(0.119)	(0.125)	(0.169)
Constant	1.350	−3.625	−1.713	3.883*	7.789
	(1.093)	(5.365)	(1.184)	(2.043)	(6.559)
Fixed effects - UF	Yes	Yes	Yes	Yes	Yes
Fixed effects - Activities	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.12	0.2	0.14	0.09	0.11
Observations	326,389	35,459	208,417	69,450	13,063
Log Likelihood	−59,414.990	−2,872.086	−33,772.930	−17,178.350	−4,019.459
Akaike Inf. Crit.	119,082.000	5,974.172	67,795.860	34,606.700	8,260.919

Note: *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 2007 and 2012 and final (t+5) 2012 and 2017.

Table 21 – Exit of activities - (2007-2017)

	Logit - Dependent variable:				
			Exit		
	General Model	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)
Relatedness Density	−8.292***	−14.051**	−11.729***	−11.332***	−12.240***
	(0.554)	(6.874)	(1.452)	(1.203)	(1.492)
PCI	2.106***	9.237***	3.443***	0.505*	−0.531
	(0.269)	(2.040)	(0.474)	(0.284)	(0.415)
Diversity	0.213***	−0.205	0.316***	0.737***	0.727*
	(0.039)	(0.203)	(0.071)	(0.115)	(0.373)
Log(GDPpc)	−0.070	−0.519	−0.212***	−0.002	−0.219
	(0.055)	(0.413)	(0.079)	(0.117)	(0.345)
Log(Population)	0.141***	−0.091	0.065	0.044	0.556***
	(0.033)	(0.170)	(0.041)	(0.066)	(0.175)
Region Productivity	0.005	1.531**	0.032	0.110	0.644
	(0.103)	(0.774)	(0.164)	(0.221)	(0.638)
Region HC	0.947**	0.300	1.382***	1.180	−1.158
	(0.384)	(1.493)	(0.526)	(0.967)	(1.813)
Incentives	0.125*	0.622***	0.173*	−0.105	−0.287
	(0.071)	(0.234)	(0.089)	(0.157)	(0.381)
Credit	0.000	0.00000***	0.000*	0.000**	−0.000
	(0.000)	(0.00000)	(0.000)	(0.000)	(0.000)
Sector Size	−0.333***	−0.264	−0.370***	−0.287***	−0.213**
	(0.044)	(0.194)	(0.057)	(0.044)	(0.084)
CL	1.753***	1.641*	1.672***	2.191***	3.626***
	(0.238)	(0.853)	(0.290)	(0.249)	(0.304)
Sector HC	0.585	4.079*	0.574	0.718	0.603
	(0.571)	(2.254)	(0.807)	(0.562)	(0.699)
Sector Productivity	0.083	−0.855*	0.092	0.163	−0.141
	(0.153)	(0.516)	(0.210)	(0.158)	(0.226)
Constant	−4.042***	−5.075	−3.121*	−6.539***	−14.457***
	(1.146)	(6.278)	(1.690)	(1.732)	(5.195)
Fixed effects - UF	Yes	Yes	Yes	Yes	Yes
Fixed effects - Activities	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.08	0.2	0.11	0.07	0.08
Observations	46,913	1,336	23,726	16,182	5,669
Log Likelihood	−25,993.630	−674.844	−12,993.650	−8,942.670	−2,847.437
Akaike Inf. Crit.	52,239.260	1,533.689	26,237.310	18,135.340	5,916.875

Note: *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 2007 and 2012 and final (t+5) 2012 and 2017.

ANNEX B – Chapter 3

Table 22 – Complex Employment Multiplier over Non-complex Employment - Brazil

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Complex emp. variation	1.757***	1.711***	2.246***	2.360***	1.878***	1.866***
	(0.351)	(0.371)	(0.326)	(0.318)	(0.323)	(0.334)
Skilled emp. share		189.349		97.949		167.492
		(137.268)		(93.460)		(109.712)
Average salary		−16.257		4.288		−11.344
		(11.647)		(25.084)		(10.863)
Relatedness		954.090***		98.842		749.570*
		(352.331)		(1,010.238)		(404.533)
Population		−0.001		−0.005*		−0.002
		(0.006)		(0.003)		(0.002)
Constant	6,635.350***	−1,862.114	5,155.273***	2,346.926	6,269.919***	−855.585
	(1,082.038)	(1,457.763)	(818.358)	(3,602.602)	(831.705)	(1,287.591)
Observations	1,116	1,116	1,116	1,116	1,116	1,116
R^2	0.705	0.717	0.703	0.705	0.705	0.716
Adjusted R^2	0.705	0.716	0.702	0.703	0.705	0.715
Residual Std. Error	16,806.120	16,498.140				
F Statistic	1,332.936***	468.713***	2,333.715***	2,455.329***	2,860.256***	3,063.636***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 23 – Complex Employment Multiplier over Non-complex Employment - Low complexity regions

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Complex emp. variation	−2.260	−1.904	1.948	1.333	−2.712	−2.450
	(2.014)	(1.852)	(4.661)	(4.613)	(2.426)	(1.932)
Skilled emp. share		−10.482		−13.708		−9.938
		(13.905)		(12.546)		(11.307)
Average salary		−21.962**		−24.147***		−21.593***
		(9.010)		(7.733)		(7.669)
Relatedness		646.410***		740.381***		630.549***
		(241.171)		(252.101)		(222.223)
Population		0.010***		0.009***		0.010***
		(0.002)		(0.002)		(0.002)
Constant	1,755.006***	313.579	1,616.346***	174.052	1,769.886***	337.129
	(353.117)	(454.079)	(430.854)	(373.606)	(365.433)	(324.723)
Observations	84	84	84	84	84	84
R^2	0.131	0.421	0.080	0.391	0.130	0.420
Adjusted R^2	0.109	0.376	0.057	0.343	0.109	0.375
Residual Std. Error	1,677.151	1,403.568				
F Statistic	6.080***	9.336***	10.077***	51.858***	12.499***	56.520***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 24 – Complex Employment Multiplier over Non-complex Employment - Medium-Low complexity regions

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Complex emp. variation	1.736**	0.826	22.172	19.534	2.202***	1.146*
	(0.780)	(0.613)	(20.851)	(18.963)	(0.794)	(0.602)
Skilled emp. share		29.340		−68.261		27.674
		(29.461)		(84.695)		(21.780)
Average salary		−8.927***		−15.880		−9.046**
		(3.312)		(13.896)		(3.586)
Relatedness		535.126***		194.948		529.319***
		(110.584)		(474.058)		(117.959)
Population		0.009***		−0.006		0.009***
		(0.002)		(0.020)		(0.002)
Constant	3,854.759***	−781.069	−1,868.199	2,668.596	3,724.303***	−722.186
	(298.359)	(785.104)	(5,800.894)	(3,274.697)	(286.111)	(639.169)
Observations	678	678	678	678	678	678
R^2	0.249	0.392	0.084	0.064	0.247	0.391
Adjusted R^2	0.247	0.387	0.081	0.056	0.244	0.386
Residual Std. Error	4,095.395	3,695.021				
F Statistic	112.050***	72.250***	25.246***	52.477***	244.952***	440.894***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 25 – Complex Employment Multiplier over Non-complex Employment - Medium-High complexity regions

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Complex emp. variation	1.952***	1.860***	3.249***	3.504***	2.049***	1.984***
	(0.293)	(0.320)	(0.316)	(0.439)	(0.291)	(0.339)
Skilled emp. share		46.610		−156.636		31.177
		(195.341)		(200.930)		(188.676)
Average salary		−11.811***		−3.768		−11.200***
		(4.524)		(4.974)		(4.149)
Relatedness		580.041***		359.693*		563.309***
		(203.150)		(197.803)		(178.600)
Population		0.002		−0.009		0.002
		(0.004)		(0.007)		(0.004)
Constant	9,662.240***	4,004.793	5,628.124***	6,168.414*	9,362.184***	4,169.087
	(936.456)	(3,719.961)	(1,005.297)	(3,589.663)	(902.486)	(3,580.855)
Observations	286	286	286	286	286	286
R^2	0.653	0.668	0.612	0.607	0.652	0.667
Adjusted R^2	0.650	0.661	0.610	0.599	0.650	0.660
Residual Std. Error	9,217.099	9,082.166				
F Statistic	266.085***	93.428***	452.372***	457.337***	546.452***	573.602***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 26 – Complex Employment Multiplier over Non-complex Employment - High complexity regions

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Complex emp. variation	1.564***	1.514***	2.178***	2.393***	1.713***	1.717***
	(0.391)	(0.374)	(0.394)	(0.447)	(0.365)	(0.342)
Skilled emp. share		3,191.032		−152.729		2,415.835
		(2,482.336)		(2,685.871)		(2,201.685)
Average salary		−58.344		48.504		−33.573
		(55.506)		(107.614)		(40.620)
Relatedness		1,462.201		−2,496.765		544.379
		(2,015.385)		(4,649.433)		(1,904.817)
Population		0.001		−0.008		−0.001
		(0.009)		(0.007)		(0.003)
Constant	31,845.440**	−4,265.097	11,092.800	40,474.100	26,828.230***	6,106.969
	(13,686.980)	(30,416.180)	(12,143.970)	(53,352.000)	(9,957.823)	(31,554.320)
Observations	68	68	68	68	68	68
R^2	0.743	0.755	0.724	0.717	0.741	0.752
Adjusted R^2	0.735	0.731	0.716	0.689	0.733	0.728
Residual Std. Error	61,867.360	62,336.340				
F Statistic	93.816***	31.307***	155.795***	148.934***	199.652***	200.799***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 27 – Non-complex Employment Multiplier over Complex Employment - Brazil

	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Non-complex emp. variation	0.390***	0.379***	0.446***	0.428***	0.395***	0.375***
	(0.050)	(0.040)	(0.057)	(0.047)	(0.054)	(0.046)
Skilled emp. share		−22.396		−43.374		−20.723
		(55.198)		(44.677)		(42.455)
Average salary		−4.943		−1.510		−5.216
		(10.222)		(8.558)		(8.716)
Relatedness		100.566		−55.863		113.039
		(404.133)		(339.486)		(348.056)
Population		0.003		0.002*		0.003**
		(0.002)		(0.001)		(0.001)
Constant	−1,630.818***	−1,569.569	−2,302.735***	−937.847	−1,693.986***	−1,619.941
	(443.129)	(1,198.809)	(526.413)	(1,080.785)	(485.105)	(1,151.255)
Observations	1,116	1,116	1,116	1,116	1,116	1,116
R^2	0.692	0.704	0.692	0.703	0.692	0.704
Adjusted R^2	0.691	0.703	0.691	0.701	0.691	0.703
Residual Std. Error	7,914.573	7,767.226				
F Statistic	1,249.502***	440.224***	2,502.499***	2,689.932***	2,451.324***	2,579.722***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 28 – Non-complex Employment Multiplier over Complex Employment - Low complexity regions

	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Non-complex emp. variation	−0.008*	−0.010*	−0.010**	0.001	−0.009**	−0.011***
	(0.004)	(0.006)	(0.005)	(0.010)	(0.004)	(0.004)
Skilled emp. share		0.879		1.007		0.861
		(0.926)		(0.778)		(0.712)
Average salary		0.453		0.694		0.420
		(0.346)		(0.494)		(0.347)
Relatedness		−22.334		−29.590		−21.320
		(16.626)		(24.364)		(19.428)
Population		0.0002		0.0001		0.0002
		(0.0002)		(0.0002)		(0.0002)
Constant	46.523*	45.315	49.623**	42.921	47.449**	45.650
	(24.678)	(34.691)	(23.105)	(41.260)	(23.405)	(43.501)
Observations	84	84	84	84	84	84
R^2	0.026	0.090	0.026	0.071	0.026	0.090
Adjusted R^2	0.002	0.019	0.002	−0.002	0.002	0.019
Residual Std. Error	100.244	99.376				
F Statistic	1.095	1.275	1.546	6.097	2.324	8.015

Note: *p<0.1; **p<0.05; ***p<0.01

Table 29 – Non-complex Employment Multiplier over Complex Employment - Medium-Low complexity regions

	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Non-complex emp. variation	0.036***	0.020*	0.052**	0.046	0.042***	0.025**
	(0.010)	(0.011)	(0.025)	(0.031)	(0.011)	(0.011)
Skilled emp. share		4.560		3.659		4.377
		(4.003)		(3.693)		(3.797)
Average salary		0.540		0.771		0.587
		(0.545)		(0.590)		(0.572)
Relatedness		7.441		−7.285		4.447
		(16.630)		(20.488)		(18.427)
Population		0.001**		0.0004		0.001
		(0.0003)		(0.001)		(0.0004)
Constant	124.870***	−166.171**	56.149	−141.186**	98.799**	−161.092**
	(40.343)	(74.941)	(108.173)	(69.917)	(45.693)	(69.605)
Observations	678	678	678	678	678	678
R^2	0.085	0.150	0.083	0.133	0.084	0.149
Adjusted R^2	0.082	0.142	0.080	0.125	0.082	0.141
Residual Std. Error	587.623	568.035				
F Statistic	31.198***	19.688***	49.188***	120.827***	77.498***	124.621***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 30 – Non-complex Employment Multiplier over Complex Employment - Medium-High complexity regions

	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Non-complex emp. variation	0.223***	0.185***	0.285***	0.277***	0.241***	0.189***
	(0.038)	(0.033)	(0.036)	(0.041)	(0.033)	(0.039)
Skilled emp. share		72.399		47.026		71.237
		(61.732)		(56.360)		(47.905)
Average salary		−1.019		0.900		−0.931
		(1.499)		(1.391)		(1.462)
Relatedness		−19.585		−95.689*		−23.071
		(52.492)		(53.969)		(54.235)
Population		0.004**		0.003		0.004***
		(0.002)		(0.002)		(0.001)
Constant	−395.127	−1,604.384*	−1,366.347***	−1,747.367*	−687.182	−1,610.934**
	(491.399)	(974.305)	(480.487)	(1,037.250)	(452.373)	(789.383)
Observations	286	286	286	286	286	286
R^2	0.505	0.587	0.500	0.564	0.504	0.586
Adjusted R^2	0.502	0.578	0.497	0.554	0.501	0.578
Residual Std. Error	3,114.047	2,866.343				
F Statistic	144.359***	65.967***	275.378***	400.046***	313.411***	397.771***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 31 – Non-complex Employment Multiplier over Complex Employment - High complexity regions

	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Non-complex emp. variation	0.397***	0.367***	0.470***	0.429***	0.406***	0.363***
	(0.061)	(0.062)	(0.066)	(0.057)	(0.068)	(0.064)
Skilled emp. share		519.451		−32.061		551.618
		(973.918)		(935.670)		(892.323)
Average salary		−32.608		−17.674		−33.480
		(42.528)		(34.549)		(33.583)
Relatedness		1,464.970		954.748		1,494.728
		(1,973.919)		(1,547.434)		(1,497.905)
Population		0.004		0.003		0.004*
		(0.003)		(0.002)		(0.003)
Constant	180.175	−21,054.210	−6,057.145	−16,044.010	−579.861	−21,346.430
	(3,809.242)	(19,788.340)	(3,942.946)	(18,464.360)	(4,398.631)	(17,987.960)
Observations	68	68	68	68	68	68
R^2	0.710	0.736	0.706	0.732	0.710	0.736
Adjusted R^2	0.702	0.710	0.697	0.705	0.701	0.710
Residual Std. Error	31,160.470	30,695.850				
F Statistic	79.732***	28.385***	157.040***	172.192***	159.020***	168.067***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 32 – Complex Employment Multiplier over Non-complex Employment - Brazil - Other classification by PCI

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Complex emp. variation	0.999***	1.028***	1.066***	1.176***	0.997***	1.064***
	(0.105)	(0.097)	(0.066)	(0.039)	(0.075)	(0.063)
Skilled emp. share		96.728		61.921		88.142
		(73.833)		(61.693)		(65.796)
Average salary		−8.213		−1.327		−6.515
		(6.676)		(11.352)		(7.638)
Relatedness		559.590**		268.539		487.799*
		(232.639)		(433.044)		(269.444)
Population		−0.005*		−0.006*		−0.005*
		(0.003)		(0.003)		(0.003)
Constant	4,312.912***	365.142	3,955.080***	1,818.492	4,322.209***	723.624
	(550.285)	(996.362)	(318.825)	(1,598.182)	(367.036)	(984.612)
Observations	1,116	1,116	1,116	1,116	1,116	1,116
R^2	0.803	0.841	0.803	0.838	0.803	0.841
Adjusted R^2	0.803	0.840	0.803	0.837	0.803	0.840
Residual Std. Error	10,647.080	9,597.411				
F Statistic	2,272.685***	975.795***	3,935.483***	5,287.281***	4,377.301***	6,049.851***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 33 – Complex Employment Multiplier over Non-complex Employment - Low complexity regions - Other classification by PCI

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Complex emp. variation	0.144	−0.593	−1.986	−1.679	0.527	−0.505
	(1.313)	(1.072)	(11.254)	(8.810)	(1.503)	(0.782)
Skilled emp. share		−11.781		−10.719		−11.867
		(14.355)		(12.750)		(11.775)
Average salary		−22.816**		−22.166***		−22.869***
		(8.863)		(7.452)		(8.138)
Relatedness		679.773***		660.766***		681.321***
		(240.649)		(211.839)		(255.261)
Population		0.010***		0.010***		0.010***
		(0.002)		(0.003)		(0.002)
Constant	1,641.155***	261.404	1,775.822**	296.669	1,616.955***	258.533
	(338.490)	(430.491)	(879.293)	(383.085)	(342.091)	(306.152)
Observations	84	84	84	84	84	84
R^2	0.114	0.408	0.097	0.404	0.113	0.408
Adjusted R^2	0.092	0.362	0.075	0.357	0.091	0.362
Residual Std. Error	1,676.910 (df = 81)	1,405.747				
F Statistic	5.202***	8.845***	10.228***	52.543***	10.479***	52.999***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 34 – Complex Employment Multiplier over Non-complex Employment - Medium-Low complexity regions - Other classification by PCI

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Complex emp. variation	1.799***	1.407***	3.341***	3.072***	2.001***	1.633***
	(0.363)	(0.368)	(0.464)	(0.540)	(0.376)	(0.397)
Skilled emp. share		30.566		24.826		29.787
		(25.513)		(24.332)		(20.702)
Average salary		−9.159***		−9.789***		−9.245***
		(2.942)		(3.603)		(3.326)
Relatedness		416.782***		312.004**		402.560***
		(99.839)		(123.299)		(104.846)
Population		0.005***		0.001		0.004**
		(0.002)		(0.002)		(0.002)
Constant	2,577.400***	−200.015	1,451.190***	434.527	2,429.568***	−113.885
	(260.111)	(691.658)	(344.099)	(581.341)	(228.005)	(535.392)
Observations	678	678	678	678	678	678
R^2	0.390	0.446	0.357	0.391	0.389	0.445
Adjusted R^2	0.389	0.441	0.355	0.386	0.388	0.440
Residual Std. Error	3,365.879	3,217.535				
F Statistic	216.220***	90.152***	327.642***	456.230***	465.665***	563.294***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 35 – Complex Employment Multiplier over Non-complex Employment - Medium-High complexity regions - Other classification by PCI

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Complex emp. variation	1.015***	1.103***	1.489***	1.662***	1.039***	1.157***
	(0.111)	(0.173)	(0.200)	(0.253)	(0.105)	(0.188)
Skilled emp. share		42.180		−52.903		32.934
		(164.167)		(157.954)		(157.862)
Average salary		−7.788**		−3.002		−7.322**
		(3.946)		(3.353)		(3.121)
Relatedness		457.385***		322.990**		444.316***
		(172.661)		(146.297)		(142.521)
Population		−0.004		−0.010**		−0.004
		(0.004)		(0.004)		(0.003)
Constant	6,802.463***	2,766.596	3,969.230***	3,437.298	6,659.012***	2,831.819
	(699.598)	(2,967.211)	(1,079.046)	(2,666.405)	(651.711)	(2,799.737)
Observations	286	286	286	286	286	286
R^2	0.652	0.674	0.629	0.648	0.652	0.674
Adjusted R^2	0.650	0.667	0.626	0.640	0.650	0.667
Residual Std. Error	7,752.148	7,560.655				
F Statistic	265.440***	96.105***	497.866***	531.921***	535.233***	586.689***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 36 – Complex Employment Multiplier over Non-complex Employment - High complexity regions - Other classification by PCI

	Dependent variable:					
	Non-complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Complex emp. variation	0.934***	0.977***	1.023***	1.190***	0.938***	1.027***
	(0.115)	(0.083)	(0.069)	(0.078)	(0.085)	(0.052)
Skilled emp. share		779.453		−504.135		481.939
		(1,406.949)		(1,405.941)		(1,289.456)
Average salary		−16.313		20.947		−7.677
		(33.743)		(49.254)		(32.473)
Relatedness		262.793		−1,079.397		−48.304
		(1,407.900)		(2,022.342)		(1,334.442)
Population		−0.005		−0.008		−0.005
		(0.003)		(0.005)		(0.004)
Constant	11,866.190*	15,630.890	6,963.532*	31,478.650	11,683.100***	19,304.140
	(6,359.463)	(20,850.930)	(4,210.323)	(25,900.780)	(4,207.490)	(20,549.060)
Observations	68	68	68	68	68	68
R^2	0.841	0.877	0.840	0.868	0.841	0.876
Adjusted R^2	0.836	0.864	0.835	0.855	0.836	0.864
Residual Std. Error	37,435.370	34,080.850				
F Statistic	172.342***	72.217***	304.936***	377.436***	337.955***	445.285***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 37 – Non-complex Employment Multiplier over Complex Employment - Brazil - Other classification by PCI

	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Non-complex emp. variation	0.788***	0.780***	0.904***	0.839***	0.812***	0.769***
	(0.101)	(0.068)	(0.051)	(0.024)	(0.046)	(0.044)
Skilled emp. share		−28.844		−48.895		−25.336
		(58.571)		(53.541)		(53.252)
Average salary		−2.809		0.508		−3.389
		(9.599)		(8.454)		(8.609)
Relatedness		−46.901		−199.739		−20.161
		(373.295)		(313.985)		(324.514)
Population		0.006***		0.006**		0.006**
		(0.002)		(0.003)		(0.003)
Constant	−2,262.525***	−2,228.733*	−3,380.496***	−1,653.450	−2,496.451***	−2,329.382*
	(801.776)	(1,242.802)	(339.895)	(1,139.275)	(283.250)	(1,208.741)
Observations	1,116	1,116	1,116	1,116	1,116	1,116
R^2	0.793	0.839	0.793	0.838	0.793	0.839
Adjusted R^2	0.793	0.838	0.792	0.837	0.793	0.838
Residual Std. Error	9,455.032	8,359.529				
F Statistic	2,135.419***	963.065***	4,040.601***	5,491.577***	4,332.588***	5,607.049***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 38 – Non-complex Employment Multiplier over Complex Employment - Low complexity regions - Other classification by PCI

	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Non-complex emp. variation	0.001	−0.004	0.001	0.015	0.001	−0.005
	(0.006)	(0.006)	(0.012)	(0.017)	(0.006)	(0.005)
Skilled emp. share		0.930		1.166		0.914
		(1.052)		(0.944)		(0.851)
Average salary		0.511		0.952*		0.479
		(0.421)		(0.507)		(0.396)
Relatedness		−14.881		−28.024		−13.944
		(17.828)		(24.700)		(20.179)
Population		0.0002		0.00002		0.0002
		(0.0002)		(0.0002)		(0.0002)
Constant	62.112**	33.387	61.552**	28.775	62.199**	33.716
	(25.636)	(38.033)	(28.414)	(41.702)	(24.791)	(43.706)
Observations	84	84	84	84	84	84
R^2	0.021	0.098	0.021	0.059	0.021	0.097
Adjusted R^2	−0.003	0.027	−0.003	−0.014	−0.003	0.027
Residual Std. Error	114.188	112.444				
F Statistic	0.872	1.388	1.743	8.734	1.743	8.458
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 39 – Non-complex Employment Multiplier over Complex Employment - Medium-Low complexity regions - Other classification by PCI

	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Non-complex emp. variation	0.128***	0.090***	0.205***	0.204***	0.139***	0.097***
	(0.021)	(0.018)	(0.041)	(0.050)	(0.027)	(0.021)
Skilled emp. share		0.260		−3.763		0.012
		(6.012)		(5.998)		(5.301)
Average salary		1.155		2.135**		1.216
		(0.741)		(0.990)		(0.741)
Relatedness		17.459		−39.949		13.915
		(22.577)		(35.605)		(23.775)
Population		0.002***		0.001		0.002***
		(0.0004)		(0.001)		(0.0005)
Constant	232.376***	−314.945***	−66.165	−231.276*	190.150*	−309.779***
	(74.975)	(104.668)	(160.458)	(127.266)	(97.309)	(97.237)
Observations	678	678	678	678	678	678
R^2	0.270	0.404	0.263	0.346	0.270	0.403
Adjusted R^2	0.268	0.398	0.261	0.340	0.268	0.398
Residual Std. Error	897.734	813.757				
F Statistic	124.858***	75.736***	185.997***	387.741***	280.595***	467.634***
Note:						*p<0.1; **p<0.05; ***p<0.01

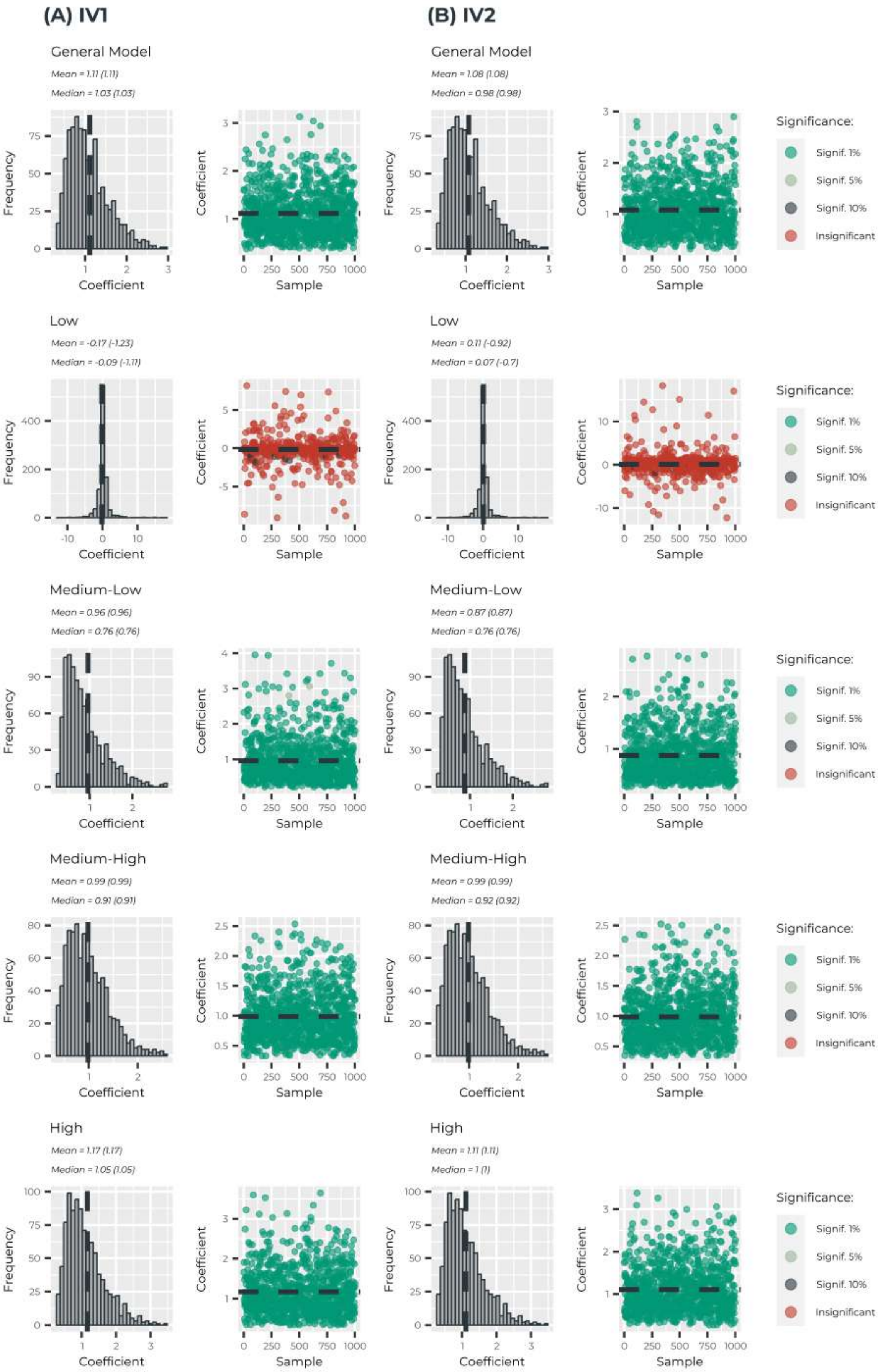
Table 40 – Non-complex Employment Multiplier over Complex Employment - Medium-High complexity regions - Other classification by PCI

	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Non-complex emp. variation	0.418***	0.345***	0.543***	0.525***	0.471***	0.359***
	(0.079)	(0.051)	(0.064)	(0.069)	(0.074)	(0.064)
Skilled emp. share		90.816		49.442		87.772
		(93.348)		(86.414)		(73.578)
Average salary		−2.615		0.487		−2.386
		(2.465)		(2.039)		(2.307)
Relatedness		−8.969		−139.039*		−18.538
		(94.449)		(79.502)		(87.174)
Population		0.009***		0.007***		0.009***
		(0.002)		(0.001)		(0.001)
Constant	599.071	−1,698.526	−1,007.929	−1,958.160	−79.365	−1,717.626
	(842.091)	(1,472.606)	(694.498)	(1,507.178)	(842.630)	(1,129.445)
Observations	286	286	286	286	286	286
R^2	0.524	0.660	0.517	0.633	0.522	0.660
Adjusted R^2	0.520	0.653	0.513	0.625	0.519	0.653
Residual Std. Error	4,972.767	4,230.193				
F Statistic	155.662***	90.382***	291.069***	515.041***	349.383***	549.036***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 41 – Non-complex Employment Multiplier over Complex Employment - High complexity regions - Other classification by PCI

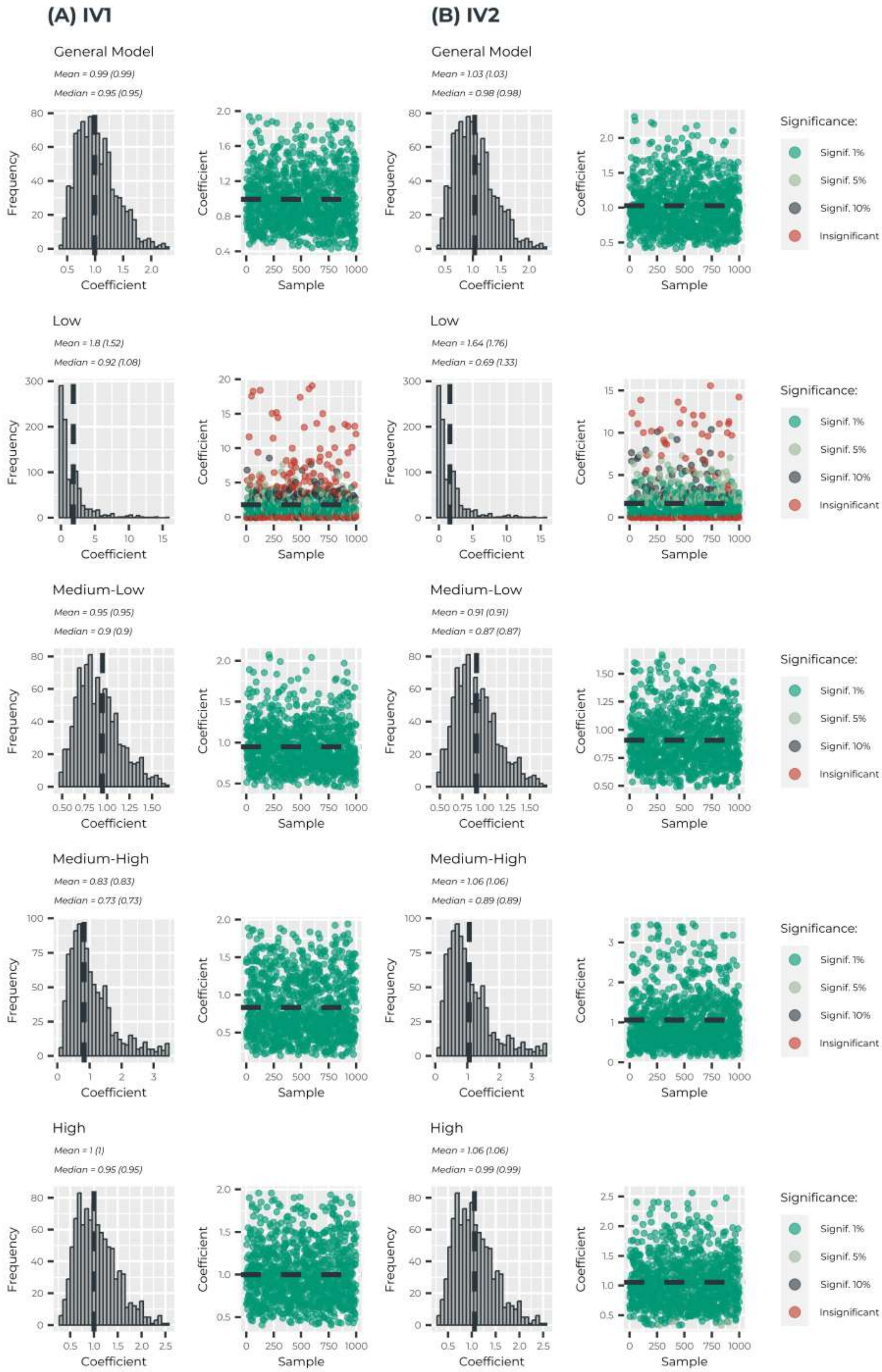
	Dependent variable:					
	Complex employment variation					
	OLS	OLS	IV_1	IV_1	IV_2	IV_2
	(1)	(2)	(3)	(4)	(5)	(6)
Non-complex emp. variation	0.816***	0.790***	0.958***	0.843***	0.845***	0.778***
	(0.122)	(0.090)	(0.044)	(0.027)	(0.038)	(0.052)
Skilled emp. share		763.086		409.629		844.896
		(1,021.613)		(1,051.104)		(1,057.935)
Average salary		−27.138		−17.211		−29.436
		(40.048)		(35.010)		(34.755)
Relatedness		1,234.200		893.755		1,312.998
		(1,818.144)		(1,471.889)		(1,454.675)
Population		0.007**		0.007*		0.007*
		(0.004)		(0.004)		(0.004)
Constant	3,425.726	−29,372.390	−5,567.473**	−26,341.440	1,572.597	−30,073.930
	(6,393.797)	(20,538.380)	(2,768.934)	(19,683.990)	(3,706.937)	(19,590.250)
Observations	68	68	68	68	68	68
R^2	0.825	0.874	0.821	0.873	0.825	0.874
Adjusted R^2	0.820	0.862	0.815	0.861	0.820	0.862
Residual Std. Error	34,986.000	30,638.590				
F Statistic	153.567***	70.705***	280.847***	403.972***	311.900***	415.117***
Note:						*p<0.1; **p<0.05; ***p<0.01

Figure 11 – Complex-Complex Multiplier - Other classification by PCI



Source: own elaboration.

Figure 12 – Non-Complex-Non-Complex Multiplier - Other classification by PCI



Source: own elaboration.