



COMPLEXIDADE ECONÔMICA E ESCALA PRODUTIVA: EVIDÊNCIAS A PARTIR DAS REGIÕES IMEDIATAS BRASILEIRAS (2006–2023)

Yasmin Nery Santos¹
Elton Eduardo Freitas²
Gustavo de Britto Rocha³

Resumo: O debate sobre desenvolvimento regional destaca o papel da especialização, da diversificação e da sofisticação produtiva nas trajetórias de crescimento e desigualdade no Brasil. Esta artigo analisa a relação entre complexidade econômica e escala produtiva nas Regiões Imediatas Brasileiras entre 2006 e 2023, investigando como a sofisticação produtiva, medida pelo Índice de Complexidade Econômica (ECI), condiciona os efeitos da especialização e da diversificação sobre os ganhos de escala. A análise utiliza dados da RAIS, DataViva, IBGE e ComexStat, cobrindo 510 Regiões Imediatas e 85 divisões da CNAE 2.0, por meio de modelos de painel com efeitos fixos. Os resultados indicam que a especialização setorial permanece como o principal motor dos ganhos de escala, inclusive em regiões mais complexas. O ECI apresenta efeito positivo, porém dependente do nível de sofisticação regional: em regiões menos complexas, ganhos marginais de complexidade ampliam a escala, enquanto em regiões altamente complexas esses efeitos se tornam modestos ou negativos, refletindo a natureza de setores intensivos em conhecimento. A diversificação tende a favorecer regiões mais sofisticadas, mas pode reduzir a eficiência em contextos intermediários.

Palavras-chave: Complexidade Econômica; Economias de Escala; Especialização; Diversificação.

ECONOMIC COMPLEXITY AND PRODUCTION SCALE: EVIDENCE FROM BRAZIL'S IMMEDIATE REGIONS (2006–2023)

Abstract: The debate on regional development highlights the role of specialization, diversification, and productive sophistication in shaping trajectories of growth and inequality in Brazil. This paper examines the relationship between economic complexity and productive scale in Brazil's Immediate Regions between 2006 and 2023, investigating how productive sophistication, measured by the Economic Complexity Index (ECI), conditions the effects of specialization and diversification on scale gains. The analysis uses data from RAIS, DataViva, IBGE, and ComexStat, covering 510 Immediate Regions and 85 divisions of CNAE 2.0, and applies panel data models with fixed effects. The results indicate that sectoral specialization remains the main driver of scale gains, even in more complex regions. The ECI has a positive effect, but its influence depends on the level of regional sophistication: in less complex regions, marginal increases in complexity significantly expand scale, whereas in highly complex regions the effects become modest or negative, reflecting the nature of knowledge-intensive sectors. Diversification tends to benefit more sophisticated regions but may reduce efficiency in intermediate contexts.

Keywords: Economic Complexity; Economies of Scale; Specialization; Diversification.

Código JEL: O4; R11

Área Temática: 1.4 Padrões de especialização produtiva e desenvolvimento

¹ Mestre em Economia na Universidade Federal de Sergipe.

² Professor Adjunto na Universidade Federal de Sergipe. E-mail: elton.freitas@academico.ufs.br

³ Professor Titular na Universidade Federal de Minas Gerais.

1. Introduction

Regional development research has long emphasized that productive structure shapes productivity, scale economies and growth trajectories. The Marshallian tradition argues that sectoral specialization, by concentrating firms and workers in space, facilitates standardized routines, reduces average costs and enables the expansion of productive scale (Marshall 1890; Krugman 1991). In contrast, the Jacobian perspective highlights that sectoral diversity supports innovation, knowledge recombination and long-run adaptability (Jacobs 1969).

More recently, the economic complexity literature has qualified this debate by arguing that diversification alone is insufficient. What matters is sophisticated diversification based on the accumulation of capabilities required to sustain advanced, non-ubiquitous activities (Hidalgo and Hausmann 2009). In this view, economic complexity can be interpreted as a form of “qualified diversification” that may alter how specialization and variety translate into performance outcomes.

This raises a question that remains largely unexplored: do complex productive structures, characterized by sophisticated diversification, reshape the classical channels through which agglomeration generates scale economies? The Brazilian case is particularly relevant because it exhibits deep territorial heterogeneity, the coexistence of industrial clusters and low-sophistication areas, and a prominent role for scale-intensive activities in manufacturing and related sectors.

This article investigates this question by examining whether economic complexity conditions the effects of specialization and diversification on productive scale across Immediate Regions in Brazil from 2006 to 2023. The central objective is to assess whether productive sophistication, measured by the Economic Complexity Index (ECI), reinforces or attenuates the mechanisms traditionally associated with scale expansion. Two hypotheses guide the analysis. The first (H1) proposes that higher regional complexity is associated with larger establishment scale, especially in low-sophistication environments. The second (H2) proposes that sectoral specialization remains the primary driver of scale, although its effect may weaken where diversification is sophisticated.

The empirical analysis relies on an annual panel covering 510 Immediate Regions and 85 divisions of the CNAE 2.0 classification over 2006–2023. Establishments and employment data are obtained from RAIS, productive complexity measures from DataViva, population and income from IBGE, and international trade from ComexStat. The econometric strategy employs fixed effects for region versus sector and year, clustered standard errors, and an extended specification that includes interaction terms and quartile-based heterogeneity with respect to ECI.

The article contributes to the literature in three main ways. First, it bridges agglomeration (specialization vs diversity), scale economies and economic complexity in a unified empirical framework. Second, it introduces complexity as a conditioning factor for the role of specialization and diversification in shaping scale, allowing for heterogeneous regional development stages. Third, it provides subnational evidence for Brazil based on administrative microdata, expanding the empirical use of complexity indicators in regional science.

The remainder of the article is organized as follows. Section 2 reviews related literature on agglomeration mechanisms, productive scale and economic complexity. Section 3 describes the data, variables and empirical strategy. Section 4 presents the main results, including interaction models and heterogeneity across complexity quartiles. Section 5 discusses the findings and their policy implications. Section 6 concludes with limitations and directions for future research.

2. Related Literature

2.1. Agglomeration economies: specialization versus diversification

A core proposition in regional science is that the spatial concentration of economic activity shapes productivity, structural change, and long-run development through agglomeration economies. In the classical marshallian account, industrial districts generate external economies because co-located firms benefit from specialized labor pools, shared input suppliers, and localized knowledge diffusion (Marshall, 1890). These localization economies imply that sectoral specialization can raise efficiency

and facilitate the emergence of scale advantages at the plant and regional level, especially when production relies on thick markets for skills and intermediate inputs.

Later contributions provided microfoundations for these mechanisms by linking agglomeration to endogenous technological change. In Arrow's (1962) learning-by-doing framework, innovation is not an exogenous shock but an outcome of cumulative experience in production. Romer (1986) formalizes the idea that knowledge is partly non-rival and imperfectly excludable, so innovations created by one firm can spill over to others, generating increasing returns at the aggregate level. The Marshall–Arrow–Romer (MAR) view thus emphasizes within-industry externalities and predicts that specialized regional structures may be particularly conducive to productivity growth where tacit knowledge and incremental innovation are important.

In contrast, Jacobs (1969) argues that the main driver of urban growth is not specialization but diversity. Her central claim is that innovation is fostered by cross-fertilization among heterogeneous activities, as varied industries and occupations create opportunities for recombination, experimentation, and the emergence of new sectors. From this perspective, diversity supports adaptability and structural transformation, especially in environments where novelty and market discovery matter.

Empirical research has tested these competing hypotheses, often finding that the relative importance of specialization and diversity depends on sectoral characteristics, spatial scale, and the outcome variable considered. Glaeser et al. (1992), using U.S. city-industry employment data, report patterns consistent with jacobian externalities, namely that sector growth correlates positively with citywide industrial diversity and negatively with own-industry specialization, while local competition appears to support innovation. Duranton and Puga (2001) propose a dynamic reconciliation: diversified “nursery cities” incubate new activities, whereas mature industries later relocate to specialized places to exploit efficiency and scale advantages. Evidence based on plant-level and panel settings further underscores heterogeneity across contexts. Henderson (2003) measures differential productivity effects of localization and urbanization externalities using manufacturing plant data, while Rosenthal and Strange (2004) synthesize the evidence on mechanisms, spatial decay, and identification strategies. A meta-analysis by Beaudry and Schiffrerova (2009) reaches a similar conclusion: results are sensitive to sector, geography, and the metric used.

More recent work has strengthened identification and measurement, emphasizing that agglomeration benefits coexist with urban costs such as congestion and land rents, and that careful strategies are needed to isolate causal effects (Combes, Duranton and Gobillon, 2011). In Brazil, Tinoco (2003) highlights that metropolitan innovation often reflects an interaction between MAR-type specialization effects and jacobian diversity effects, rather than a strict either-or mechanism.

For the purposes of this study, this literature motivates two key points. First, specialization and diversity are not mutually exclusive, and their effects can be stage-dependent, with diversity often tied to innovation and resilience, and specialization more closely linked to efficiency. Second, the literature leaves open how the specialization–diversity debate interacts with the scale of production, particularly when regional structures differ not only in variety but also in the sophistication of their activities.

2.2. Economies of scale and their link to agglomeration

The idea that larger scale can raise efficiency has deep roots in economic theory. Smith (1776) emphasized division of labor as a source of productivity gains, while Marshall (1890) distinguished internal economies of scale, arising within firms through organization and managerial practices, from external economies, generated by the broader industrial environment. A crucial implication for regional analysis is that internal scale advantages may be enabled, constrained, or reshaped by spatial context.

In growth and trade, scale considerations were incorporated into theories of cumulative causation and increasing returns. Kaldor (1966) associated industrial expansion with dynamic increasing returns, reinforcing the role of manufacturing in productivity growth. In the New Economic Geography, Krugman (1979, 1991) formalizes how increasing returns and transport costs jointly generate spatial concentration, thereby linking scale to location decisions. Fujita, Krugman and Venables (2001) provide a canonical synthesis of these mechanisms, highlighting how market access, imperfect competition, and scale economies shape agglomeration patterns. These contributions also clarify that economies of scale

(declining average costs) and increasing returns (more-than-proportional output responses) are related but conceptually distinct, a distinction that matters when moving from theory to empirical measurement.

A central empirical question then becomes how, in practice, firms realize internal scale economies and whether agglomeration facilitates this process. Holmes and Stevens (2002) provide a seminal contribution by showing that geographically concentrated industries tend to operate with larger establishments. Their interpretation is that localization economies reduce coordination, search, and matching costs, enabling firms to operate at larger scale without proportional cost increases. This result has been replicated in other contexts. Barrios, Bertinelli and Strobl (2006) confirm similar relationships for Ireland, while Lafourcade and Mion (2007) show that industrial geography affects the distribution of establishment sizes in France. Holmes and Stevens (2014) further refine the interpretation by distinguishing standardized, mass-market plants that are typically large and concentrated from customized production plants that are smaller and more dispersed.

Related work suggests that agglomeration can raise measured productivity through multiple channels that are complementary to scale. Behrens, Duranton and Robert-Nicoud (2014) integrate agglomeration economies, firm selection, and worker sorting to explain the productivity advantage of larger cities. In such settings, scale outcomes are not merely firm decisions but part of an equilibrium in which more efficient firms and more skilled workers disproportionately locate in dense regions.

For Brazil, Soares, Santos and Freitas (2008), using RAIS data and a panel framework for microrregions, document a positive relationship between specialization and productive scale, emphasizing that omitting regional and sectoral heterogeneity may bias inference. Even though Bressan, Braga and Bressan (2008) focus on a specific sector, their evidence also reinforces the broader point that cost efficiency and performance are scale-sensitive and context-dependent.

A further extension is that the sources of scale economies may change with technological paradigms. Shapiro and Varian (1999) and Brynjolfsson and McAfee (2014) emphasize strong scale properties in information goods with near-zero marginal costs. Perez (2002) and Mazzucato (2013) highlight the role of institutions and the state in enabling large-scale, high-risk investments during technological transitions. While these contributions are not primarily spatial, they reinforce a principle that is central to regional science: scale advantages often depend on external arrangements, whether geographic, network-based, or institutional.

Taken together, the scale literature suggests a tight link between agglomeration and establishment size in many manufacturing settings, but also points to potential heterogeneity across sectors and technological regimes. This motivates a natural extension: if the productive structure becomes more sophisticated and knowledge-intensive, does the classic specialization-to-scale channel weaken, persist, or transform?

2.3. Economic complexity: sophisticated diversification, capabilities, and regional trajectories

The economic complexity approach qualifies the specialization–diversification debate by focusing on the nature of the activities present in a region. Rather than treating diversification as a simple count of sectors, complexity emphasizes that development depends on the accumulation of productive capabilities that are rare, complementary, and difficult to imitate, which in turn shape the set of feasible activities and the direction of structural change (Hidalgo and Hausmann, 2009).

The empirical foundations of this literature are grounded in network representations of production. The Product Space framework maps the proximity among products based on similarity in required capabilities, highlighting that moving into more sophisticated activities is path dependent and constrained by relatedness (Hausmann and Klinger, 2006). Building on this insight, Hidalgo et al. (2007) and Hidalgo and Hausmann (2009) introduce the Economic Complexity Index (ECI), which combines diversity and ubiquity to capture productive sophistication. Tacchella et al. (2012) propose the Fitness–Complexity approach as a nonlinear alternative aimed at addressing limitations of linear measures.

A substantial empirical literature associates complexity with income, growth, and distributional outcomes. Hartmann et al. (2016) show that more complex economies tend to display lower income inequality, suggesting that sophisticated productive structures may support broader inclusion. Gala, Rocha and Magacho (2018) similarly report that higher complexity is associated with a greater

probability of income convergence. At the country level, comparative analyses suggest that early specialization in complex products can shape development trajectories, as illustrated by the contrast between Brazil and South Korea (Britto et al., 2015) and by evidence that richer economies occupy central positions in global trade networks through complex specializations (Gala, Camargo and Freitas, 2017).

The framework is especially relevant for Latin America, where debates on deindustrialization and commodity dependence are central. Alencar et al. (2018) document stagnation in low-complexity products for several countries in the region, linking this pattern to commodity booms and industrial decline. For Brazil, Nunes (2024) reports evidence consistent with productive “de-complexification”, with increasing export shares in low-complexity commodities and declining participation of more sophisticated sectors. Quinzani and Borges (2020) interpret development as structural transformation driven by value added and innovation, where sophistication captured by complexity measures is a key mechanism.

From a regional perspective, complexity is closely tied to related diversification and capability accumulation. Regions tend to diversify into industries that are proximate to their existing knowledge base, a pattern documented by Neffke, Henning and Boschma (2011). Brazilian evidence reinforces this view, showing that relatedness and complexity shape diversification pathways and can generate asymmetric regional outcomes (Freitas, Britto and Amaral, 2024; Queiroz, Romero and Freitas, 2024). Balland et al. (2018) connect these mechanisms to smart specialization strategies, arguing that diversification and innovation are anchored in preexisting capabilities and thus display strong path dependence. More recent syntheses emphasize the broad applicability of economic complexity to industrial policy, regional development, and new challenges such as the green transition (Hidalgo, 2021; Mealy, Farmer and Teytelboym, 2019, 2022; Romero and Gramkow, 2021).

For this article, the key conceptual move is to interpret complexity as sophisticated diversification. A region may be diversified in many low-knowledge activities, or diversified in a portfolio of capability-intensive activities. These two forms of diversity are likely to imply different relationships between agglomeration and productive scale. In less complex regions, incremental increases in sophistication may reinforce traditional scale mechanisms, for example by improving supplier networks, labor matching, and the feasibility of larger plants. In highly complex regions, however, productive efficiency may become less dependent on establishment size, because knowledge-intensive activities can exhibit modular production, higher reliance on intangible inputs, and weaker dependence on large-scale plants.

This perspective highlights a gap in the existing evidence. The agglomeration literature documents when specialization or diversity supports growth, and the scale literature connects specialization to establishment size, but relatively little is known about how these relationships vary with the sophistication of the regional productive structure. By explicitly modeling economic complexity as a conditioning factor in the specialization–diversification–scale nexus, the present study contributes to integrating three research traditions, agglomeration economies, economies of scale, and economic complexity, within a unified empirical framework tailored to regional heterogeneity.

3. Data and Empirical Strategy

3.1. Data and unit of analysis

This study investigates how productive scale relates to regional productive sophistication, specialization, and diversification in Brazil over the period 2006-2023. The empirical analysis relies on an annual panel indexed by Immediate Geographic Region r , sector s , and year t . Immediate Regions are the basic spatial unit because they are designed to capture short-range functional linkages among municipalities, including commuting, service provision, and local market interactions, which are precisely the channels through which agglomeration and scale mechanisms are expected to operate (IBGE, 2017).

The dataset combines multiple official sources. Employment and establishment counts by sector and region come from RAIS. Regional economic complexity measures are obtained from DataViva, following employment-based implementations of complexity indices in the Brazilian context

(FREITAS, 2019; FREITAS et al., 2023; FREITAS; BRITTO; AMARAL, 2024). Regional macro controls, including GDP and population, are drawn from IBGE, while external openness is proxied using regional trade flows from ComexStat/IBGE.

The final sample covers 510 Immediate Regions and 85 CNAE 2.0 divisions, excluding division 84 (Public Administration, Defence and Social Security) and division 99 (International Organisations and Other Extraterritorial Institutions). These exclusions avoid distortions from non-market or administratively defined activities and ensure that the analysis focuses on private-sector productive dynamics with consistent measurement across territories and over time. Sectoral information is kept at the division level to preserve comparability, reduce volatility from very fine classifications, and retain sufficient within-region variation for panel estimation.

3.2. Variables: measurement, interpretation, and sources

Table 1 summarizes the main variables, their sources, and expected signs according to the literature. The dependent variable captures productive scale, while the main explanatory variables represent specialization, diversification, and economic complexity, complemented by controls capturing human capital, market size, income, and openness.

Productive scale is measured as the region–sector average establishment size relative to the national sectoral average:

$$Scale_{r,s,t} = \frac{\frac{Emp_{r,s,t}}{Est_{r,s,t}}}{\frac{Emp_{s,t}}{Est_{s,t}}} \quad (1)$$

where $Emp_{r,s,t}$ and $Est_{r,s,t}$ denote employment and number of establishments in sector s and region r at time t , and $Emp_{s,t}$ and $Est_{s,t}$ are the corresponding national totals for sector s at time t . Values above one indicate that the average establishment size in r - s is larger than the national mean in that sector-year, consistent with stronger realized scale at the plant level.

This measure is closely connected to the decomposition of the location quotient proposed by Holmes and Stevens (2002), which separates specialization into an establishment component and an establishment-size component. For Brazil, a similar decomposition and interpretation is employed by Soares, Santos and Freitas (2008), emphasizing its usefulness for studying heterogeneity in the specialization–scale relationship.

Regional economic complexity is measured by the Economic Complexity Index (ECI) computed from employment data in DataViva, following the methodological adaptations used for Brazilian regional complexity (FREITAS, 2019; FREITAS et al., 2023; FREITAS; BRITTO; AMARAL, 2024). The index is rooted in the original formulation by Hidalgo and Hausmann (2009), combining:

- i. Diversity, the number of activities in which a region displays revealed comparative advantage, and
- ii. Ubiquity, the number of regions that are active in the same activities.

Higher ECI values indicate regions that are diversified in less ubiquitous (more capability-intensive) activities. For comparability across time and units, the ECI is normalized to a 0–100 range using a min–max transformation.

Sectoral specialization is proxied by the location quotient (LQ):

$$LQ_{r,s,t} = \frac{\frac{Emp_{r,s,t}}{Emp_{r,t}}}{\frac{Emp_{s,t}}{Emp_t}} \quad (2)$$

where $Emp_{s,t}$ is total regional employment and Emp_t is total national employment. $LQ > 1$ indicates that sector s has a higher employment share in region r than in the national economy, a standard indicator of relative specialization used widely in regional studies of agglomeration and growth (GLAESER et al., 1992; COMBES, 2000; FREITAS, 2012; SOARES; SANTOS; FREITAS, 2008).

A known concern is the “small-region bias” in which LQs can be inflated in small economies with thin sectoral bases. To mitigate this, the analysis follows the leave-one-out correction proposed by Holmes and Stevens (2002), which reduces mechanical correlation and finite-sample distortions.

Regional diversification is measured using the Shannon entropy index, which accounts for both the number of sectors and the distribution of employment shares:

$$H_{r,t} = - \sum_{i=0}^S p_{r,s,t} \cdot \ln(p_{r,s,t}) \quad (3)$$

where $p_{r,s,t}$ is sector s 's share in total employment in region r at time t . Higher values indicate a more diversified local employment structure (PASCHOALINO; PARRÉ; CHAGAS, 2025).

Controls are included to reduce omitted-variable bias and to account for regional differences in market size, income, human capital, and external exposure. Human capital is proxied by the share of workers with completed tertiary education in region–sector employment:

$$HC_{r,s,t} = \frac{Graduates_{r,s,t}}{Emp_{r,s,t}} \quad (4)$$

following the use of educational attainment as a measure of local skill composition in urban and regional studies (GLAESER; MARÉ, 2001; FREITAS, 2012).

Trade openness is measured by the trade-to-GDP ratio:

$$Trade_{r,t} = \frac{Imp_{r,t} + Exp_{r,t}}{GDP_{r,t}} \quad (5)$$

where exports and imports are from ComexStat/IBGE and GDP is from IBGE. In addition, the model includes $\ln(GDPpc_{r,t})$ and $\ln(POP_{r,t})$, capturing income and market size. GDP per capita values are deflated using IPCA (base year 2020) to ensure intertemporal comparability.

3.3. Econometric specification and identification strategy

The empirical strategy estimates the association between productive scale and the regional productive structure using panel data models. Panel methods are appropriate because they exploit both cross-sectional variation across regions and sectors and temporal variation, while controlling for time-invariant unobserved heterogeneity (WOOLDRIDGE, 2010). The baseline specification is:

$$Scale_{r,s,t} = \beta_0 + \beta_1 ECI_{r,s,t} + \beta_2 LQ_{r,s,t} + \beta_3 H_{r,t} + \delta' X_{r,t} + \alpha_{r,s} + \gamma_t + \varepsilon_{r,s,t}$$

where $X_{r,t}$ is a vector of controls (human capital, trade openness, log GDP per capita, log population). The term $\alpha_{r,s}$ denotes region–sector fixed effects, absorbing all time-invariant factors specific to a given region-sector pair (for example, persistent locational advantages, historical industrial composition, and long-run institutional differences). Year fixed effects γ_t capture common macroeconomic shocks, national policy changes, and aggregate business cycle variation.

Standard errors are clustered at the region level, allowing for arbitrary serial correlation and cross-sector dependence within each region over time, which is consistent with the fact that shocks to local labor markets and infrastructure may affect multiple sectors simultaneously.

To test whether the effects of specialization and diversification depend on regional sophistication, the empirical strategy introduces interaction terms:

- $ECI_{r,s,t} \cdot LQ_{r,s,t}$
- $LQ_{r,s,t} \cdot H_{r,t}$
- $ECI_{r,s,t} \cdot H_{r,t}$

These interactions operationalize the article's central mechanism: economic complexity may requalify the classic specialization–scale link, and it may also determine whether diversification strengthens or weakens scale performance. In particular, the coefficient on $ECI \times LQ$ indicates whether specialization is more strongly associated with scale in less versus more complex regions, while $ECI \times H$ captures whether diversified regions benefit differently depending on their sophistication, consistent with the idea of “sophisticated diversification”.

Finally, to assess non-linearities and heterogeneous responses, the effect of specialization on scale is estimated separately by ECI quartiles. This approach complements the interaction models by allowing

the specialization–scale relationship to vary flexibly across low- and high-complexity regional regimes, providing a direct test of whether specialization remains the primary driver of scale even among sophisticated regions.

4. Results

4.1. Baseline results: descriptive evidence and direct effects

We begin with a descriptive overview of the variables used in the empirical analysis. Tables 1 and 2 reports summary statistics and the correlation matrix for the full panel of Immediate Geographic Regions, sectors and years. The dataset comprises almost 500 thousand region–sector–year observations between 2006 and 2023, which allows us to span distinct phases of the Brazilian business cycle, including the mid-2000s expansion, the 2014–2016 recession and the subsequent partial recovery.

The dependent variable, productive scale, displays a low mean of about 0.78, combined with very high dispersion, with a standard deviation of 1.60 and extreme values above 260. This pattern indicates that a small number of region–sector pairs operate with very large establishments relative to the national sectoral average, while most combinations are characterized by relatively small plants. The Economic Complexity Index (ECI) ranges from 0 to 100, with a mean around 40 and a standard deviation above 20. This confirms that only a limited group of regions reach high levels of productive sophistication, whereas a large share of the territory remains at low or intermediate levels.

Table 1 - Descriptive Statistics of the Model Variables

Variables	Observations	Mean	Standard Deviation	Min	Max
Scale	499,797	0.78	1.60	0.00	260.09
ECI	499,797	40.13	20.84	0.00	100.00
LQ	499,797	1.29	6.08	0.00	613.49
Diversification	499,797	2.85	0.40	0.68	3,66
Human Capital	499,797	10.68	16.24	0.00	100.00
Trade	499,797	6.81	12.11	0.00	200.23
GDP pc (<i>ln</i>)	499,797	10.08	0.61	8.30	12.47
Population (<i>ln</i>)	499,797	12.38	0.95	10.08	16.91

Source: Prepared by the author based on data from RAIS, DataViva, IBGE, and ComexStat.

Table 2 - Correlation of Model Variables

Variables	Scale	ECI	LQ	Diversification	Human Capital	Trade	GDP pc (<i>ln</i>)	Population (<i>ln</i>)
Scale	1.0000							
ECI	-0.0006	1.0000						
LQ	0.3832	-0.0136	1.0000					
Diversification	0.0600	0.4290	-0.0171	1.0000				
Human Capital	0.0391	0.0747	-0.0139	0.1381	1.0000			
Trade	0.0326	0.0846	0.0154	0.1117	-0.0047	1.0000		
GDP pc (<i>ln</i>)	0.0343	0.5215	-0.0138	0.4935	0.1308	0.2983	1.0000	
Population (<i>ln</i>)	0.0980	0.0064	-0.0234	0.5790	0.1103	0.0434	0.1265	1.0000

Source: Prepared by the author based on data from RAIS, DataViva, IBGE, and ComexStat.

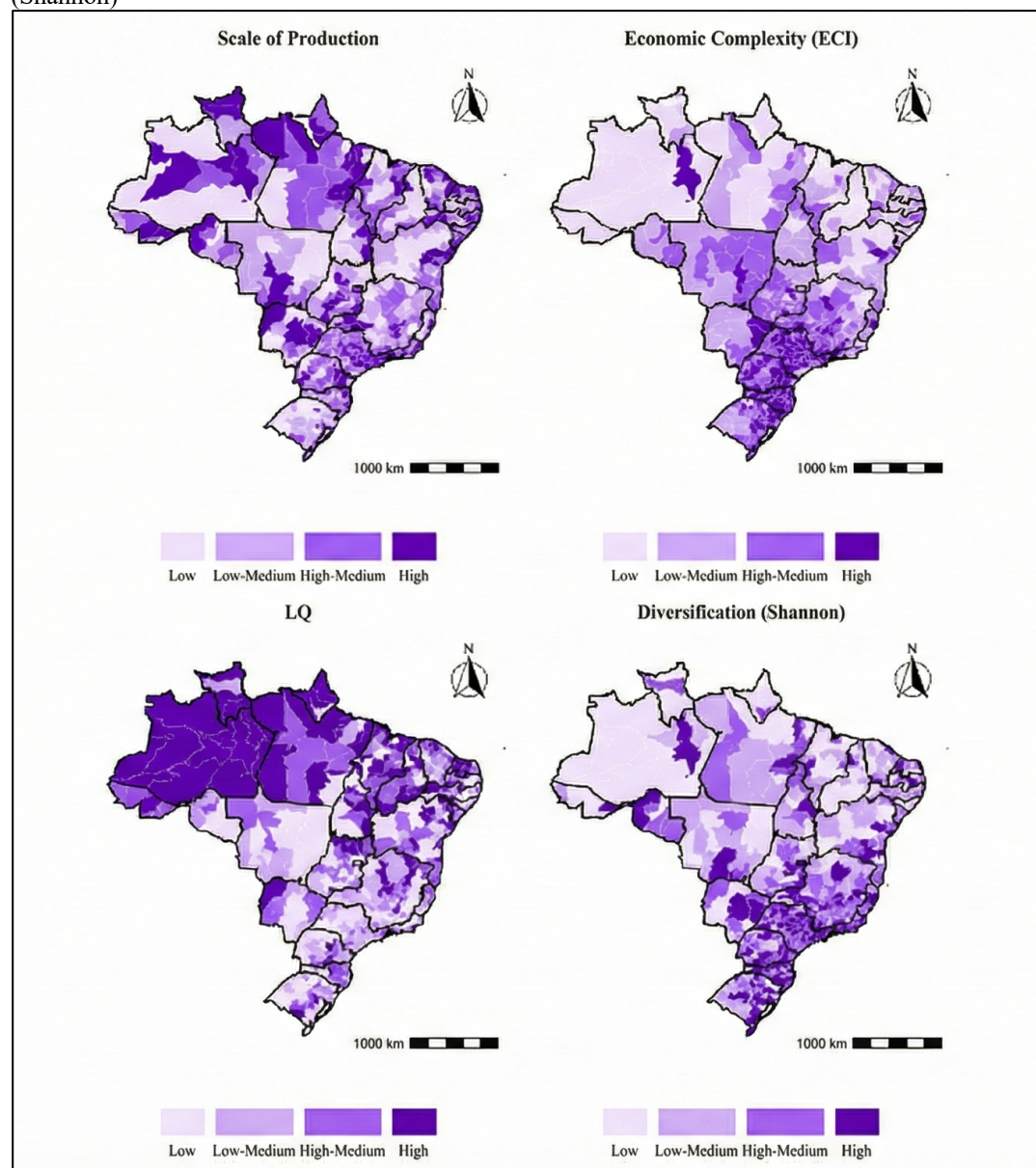
The location quotient (LQ) has extremely wide variation, from zero to values above 600, reflecting cases of very strong regional specialization as well as complete absence of a given sector. Diversification, measured by the Shannon entropy index, shows a mean around 2.85 and moderate dispersion, indicating that many regions exhibit an intermediate degree of sectoral variety, while others remain highly concentrated. Human capital is also very unevenly distributed: the share of workers with tertiary education averages about 11 percent but ranges from virtually zero to 100 percent, revealing stark differences in skill composition across regions. Trade openness, measured as the trade-to-GDP ratio, is close to 7 percent on average, yet it reaches values above 200 percent in some regions and

remains at zero in others. Finally, the distributions of log GDP per capita and log population reflect the well-known contrast between large, high-income metropolitan areas and small, low-income regions.

The correlation matrix in Table 2 does not suggest problematic multicollinearity. The dependent variable, productive scale, is moderately correlated with LQ (around 0.38), indicating that more specialized region–sector pairs tend to operate at larger scale. Its correlations with ECI and diversification are close to zero or weak, which is helpful for identification. ECI is strongly correlated with log GDP per capita and diversification, which is consistent with the idea that more complex regions are both richer and more diversified. Diversification, in turn, is strongly associated with population size, which reflects the tendency of large urban areas to support more varied productive structures. Human capital and trade openness show only modest pairwise correlations with other variables, so their effects are more likely to be revealed in the multivariate regressions.

Spatial patterns provide further insight. Figure 1 maps the regional averages of the main variables over 2006–2023. Productive scale is clearly concentrated in the Centre–South corridor, especially in the interior of São Paulo, Paraná, Santa Catarina and Rio Grande do Sul, as well as in parts of the Northeast and in agro-industrial frontiers in the West of Bahia and the southern fringe of the Amazon. These are areas where large industrial plants and capital-intensive agro-industries are located, consistent with the traditional link between concentration, specialization and scale economies.

Figure 1 – Average of the Main Variables in the Econometric Model (2006–2023): Scale, ECI, LQ and Diversification (Shannon)



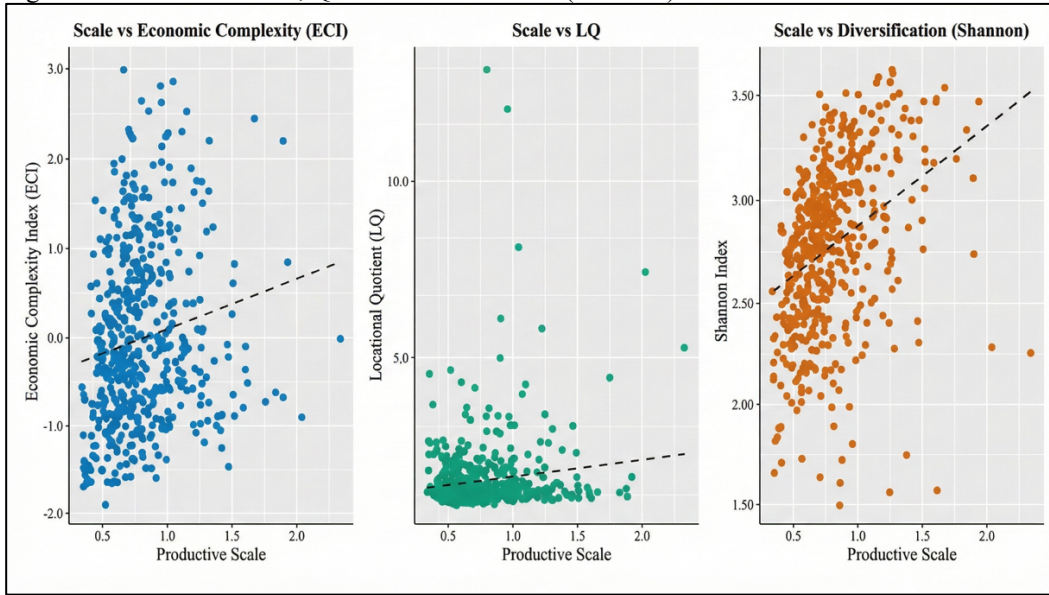
Source: Prepared by the author based on data from RAIS, DataViva, IBGE, and ComexStat.

The ECI map also shows a strong concentration in the South and Southeast, with additional complex nodes along the Northeastern coast, particularly in Bahia, Pernambuco and Ceará, where more sophisticated manufacturing and service activities are present. In contrast, much of the North and inland Northeast displays low complexity, reflecting dependence on primary sectors and low-tech activities. This pattern is in line with the view that productive capabilities accumulate unevenly across space and underpin complex diversification paths.

The LQ map highlights very high specialization in parts of the North, notably in Amazonas and Pará, driven by mining and the Manaus industrial pole. However, when LQ is compared with the Shannon index, it becomes clear that these highly specialized regions are much less diversified and therefore more vulnerable to sector-specific shocks. By contrast, many regions in the South, Southeast and coastal Northeast combine high scale, high ECI and high diversification, which suggests more resilient, sophisticated and adaptable productive systems.

Figure 2 presents scatter plots of productive scale against ECI, LQ and diversification. The relationship between scale and ECI is positive but relatively mild and noisy, which suggests that complexity is associated with larger plants on average but that other structural factors also play an important role. The relationship between scale and LQ is more dispersed and asymmetric, with a dense cluster of observations at low LQ and a long tail of highly specialized, high-scale outliers, especially in mining and the Manaus Free Trade Zone. The relationship between scale and diversification is more clearly positive, indicating that diversified regions tend to host larger establishments, in line with the jacobian view that variety stimulates complementarities and innovation.

Figure 2 - Scale versus ECI, QL and Diversification (Shannon)



Source: Prepared by the author based on data from RAIS, DataViva, IBGE, and ComexStat.

These descriptive patterns motivate the econometric analysis. They suggest that the scale of production is shaped jointly by specialization, diversification and complexity, and that these channels are likely to operate differently across types of regions.

Turning to the baseline panel regressions, Table 3 reports fixed-effects estimates for models with productive scale as the dependent variable. The first set of columns presents specifications without interactions. Across all versions, the coefficient on LQ is positive, large and statistically significant. The magnitude is around 0.10–0.11, meaning that a one-unit increase in LQ is associated with an increase of roughly 0.10 in the relative establishment size, controlling for ECI, diversification, human capital, trade openness, log GDP per capita, log population and fixed effects. This confirms that sectoral specialization remains a central driver of scale, consistent with MAR-type agglomeration mechanisms.

The coefficient on ECI is positive and significant in all baseline models, typically in the range of 0.0007–0.005 depending on the inclusion of fixed effects. Given that ECI has been normalized to the 0–100 range, these numbers imply that moving from a very low-complexity region to a highly complex

one is associated with nontrivial increases in average establishment size. However, the ECI effect is smaller than that of LQ and less stable across specifications, which suggests that complexity is important but operates as a more subtle conditioning factor.

The behaviour of diversification is more complex. In simpler specifications, diversification appears negatively associated with scale, with coefficients around -0.04 to -0.06 , which is consistent with the idea that, in some contexts, spreading resources across many activities can weaken agglomeration benefits. Once region, sector and year fixed effects are introduced, however, the coefficient turns positive and significant, with values around 0.11 to 0.12 . This indicates that, after controlling for unobserved heterogeneity, diversified regions tend to support larger establishments, which aligns with the view that variety can enhance complementarities, innovation and resilience.

Table 3 - Results of the estimated models for the dependent variable Productive Scale of the Immediate Brazilian Regions (2006–2023)

	(1)	(2)	(3)	(4)	(5)	(6)
ECI	0.00189*** (0.000249)	0.00493*** (0.000439)			0.000757*** (0.000146)	0.000323** (0.000146)
LQ			0.108*** (0.00493)	0.109*** (0.00495)	0.108*** (0.00493)	0.109*** (0.00495)
Diversification			-0.0601*** (0.0160)	0.118*** (0.0148)	-0.0430** (0.0174)	0.113*** (0.0153)
Human Capital			0.00304*** (0.000130)	0.00481*** (0.000244)	0.00304*** (0.000129)	0.00482*** (0.000244)
Trade Openness			0.00214*** (0.000233)	0.00144*** (0.000234)	0.00207*** (0.000232)	0.00148*** (0.000237)
Log GDP pc			0.0684*** (0.00535)	0.205*** (0.00705)	0.0788*** (0.00552)	0.201*** (0.00733)
Log Population			0.177*** (0.00360)	0.0912*** (0.00361)	0.173*** (0.00393)	0.0924*** (0.00370)
Constant	0.739*** (0.0106)	1.981*** (0.395)	-2.111*** (0.0569)	-2.376*** (0.0715)	-2.178*** (0.0576)	-2.347*** (0.0731)
Year FE	No	Yes	No	Yes	No	Yes
State FE	No	Yes	No	Yes	No	Yes
Sector FE	No	Yes	No	Yes	No	Yes
R2	0.000133	0.0177	0.160	0.186	0.160	0.186
Adj R2	0.000132	0.0174	0.160	0.186	0.160	0.186
Observations	565,339	565,339	434,691	434,691	434,691	434,691
F-test	57.50	192.1	1,538.5	351.1	1,322.5	348.4

Source: Prepared by the author based on data from RAIS, DataViva, IBGE, and ComexStat.

Note: Standard errors are in parentheses * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

The control variables behave as expected. Human capital has a positive and significant coefficient in all models, indicating that regions with higher shares of tertiary-educated workers host larger plants. Trade openness is also positively associated with scale, suggesting that integration into external markets supports the expansion of capacity. Log GDP per capita and log population show positive and highly significant effects, reinforcing the idea that richer and larger markets provide favourable conditions for agglomeration and scale.

The adjusted R^2 increases substantially as controls and fixed effects are added, reaching values around 0.16 to 0.19 in the most complete specifications. This indicates that the combination of specialization, complexity, diversification and regional characteristics explains a meaningful share of the variation in productive scale across region–sector pairs and over time.

In sum, the baseline results show that specialization is the strongest direct correlate of productive scale, while complexity and diversification also contribute, but in ways that are more sensitive to model specification and the treatment of heterogeneity. This motivates the next step, which explicitly investigates how complexity conditions the effects of specialization and diversification.

4.2. Interaction effects: specialization, diversification and complexity

The second panel of Table 4 presents models with interaction terms between ECI, specialization and diversification. These specifications are designed to test whether the impact of specialization and diversification on scale depends on the level of regional complexity.

Table 4 - Results of the estimated models for the dependent variable Productive Scale of the Immediate Brazilian Regions with Interaction Terms (2006–2023)

	(1)	(2)	(3)	(4)
ECI	0.00116*** (0.000271)	0.00733*** (0.00147)	0.000337** (0.000146)	0.00870*** (0.00140)
LQ	0.135*** (0.0109)	0.109*** (0.00495)	0.161*** (0.0445)	0.119*** (0.0457)
Diversification	0.116*** (0.0151)	0.200*** (0.0291)	0.137*** (0.0170)	0.201*** (0.0251)
Human Capital	0.00478*** (0.000245)	0.00480*** (0.000244)	0.00483*** (0.000244)	0.00475*** (0.000244)
Trade Openness	0.00148*** (0.000236)	0.00155*** (0.000237)	0.00147*** (0.000237)	0.00156*** (0.000236)
Log GDP pc	0.198*** (0.00735)	0.198*** (0.00745)	0.202*** (0.00749)	0.195*** (0.00766)
Log Population	0.0935*** (0.00375)	0.0940*** (0.00361)	0.0919*** (0.00358)	0.0955*** (0.00344)
ECI x LQ	-0.000626*** (0.000203)			-0.000687*** (0.000204)
ECI x Diversification		-0.00228*** (0.000472)		-0.00243*** (0.000439)
LQ x Diversification			-0.0188 (0.0154)	0.00646 (0.0166)
Constant	-2.379*** (0.0758)	-2.609*** (0.0907)	-2.425*** (0.0978)	-2.635*** (0.105)
Year FE	Sim	Sim	Sim	Sim
State FE	Sim	Sim	Sim	Sim
Sector FE	Sim	Sim	Sim	Sim
R2	0.189	0.186	0.187	0.189
Adj R2	0.189	0.186	0.187	0.189
Observations	434,691	434,691	434,691	434,691
F-test	345.8	347.9	346.2	346.4

Source: Prepared by the author based on data from RAIS, DataViva, IBGE, and ComexStat.

Note: Standard errors are in parentheses * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

The most robust and economically meaningful result concerns the interaction between ECI and LQ. The coefficient on $ECI \times LQ$ is negative and statistically significant, with values around -0.0006 to -0.0007 , while the main effect of LQ remains strongly positive. This pattern implies that specialization increases productive scale in all regions, but the marginal effect of specialization becomes weaker as complexity rises. In low-complexity regions, an additional unit of LQ is associated with sizeable scale gains, consistent with marshallian mechanisms that rely on spatial concentration to exploit internal economies of scale. In highly complex regions, by contrast, the same increase in specialization has a much smaller impact, and in the upper range of ECI values the net marginal effect of specialization on scale approaches zero.

This finding is in line with the article's central hypothesis that economic complexity requalifies the classic specialization–scale channel. In regions where the productive structure is still dominated by relatively simple activities, concentrating resources in a few sectors is an effective way to generate scale. In regions where the structure is already dense in capabilities, many activities are knowledge intensive and modular, so they do not require very large plants to remain competitive.

The interaction between ECI and diversification is also negative and statistically significant in all specifications. Coefficients for $\text{ECI} \times \text{diversification}$ lie around -0.0023 to -0.0024 . This means that the positive association between diversification and scale tends to be stronger in less complex regions and weaker in more complex ones. In other words, diversification is most scale enhancing when it helps low-complexity economies broaden their productive base and build new complementarities. In high-complexity regions, by contrast, the marginal gains from spreading activity across additional sectors are reduced, since the underlying capability base and network of linkages are already dense.

The interaction between LQ and diversification is small and statistically insignificant in all models. This suggests that, in the Brazilian case, it is the level of complexity rather than the combination of specialization and diversification per se that conditions how these channels translate into larger establishment sizes.

Taken together, the interaction results show that complexity acts as a modulator of both specialization and diversification. It does not simply add another direct effect on scale. Instead, it reshapes how traditional agglomeration channels operate, reinforcing them in low-complexity environments and attenuating them in highly complex ones.

4.3. Heterogeneity by complexity levels: ECI quartiles

To explore non-linearities in more detail, Table 3 reports separate regressions for four groups of regions defined by quartiles of the ECI distribution: Low, Low-medium, High-medium, and High complexity. All specifications include the same set of controls and fixed effects as in the baseline models.

In Low complexity regions (first quartile), the coefficient on ECI is positive and significant across most specifications, with values ranging approximately from 0.003 to 0.011 depending on controls and fixed effects. This confirms that even modest increases in sophistication are associated with higher productive scale when the regional capability base is initially shallow. The coefficient on LQ is also positive and significant, around 0.10 to 0.11, indicating that specialization remains a strong driver of scale. Diversification exhibits mixed effects in this group: in some models it is positive and significant, whereas in others it is not statistically different from zero. Human capital, trade openness, log GDP per capita and log population all display robust positive effects. The adjusted R^2 of the most complete models reaches around 0.14, which is high given the size and heterogeneity of the sample.

In Low-medium complexity regions (second quartile), the results for ECI are considerably more fragile. In the simplest specifications, the coefficient on ECI is slightly negative and not significant. Only when additional controls are included does ECI become positive and significant; once full fixed effects are added, the coefficient loses robustness and can even turn negative. This indicates that, at intermediate levels of sophistication, complexity yields scale gains only when combined with other favourable conditions, such as higher skill endowments and stronger external integration. The coefficient on LQ remains positive and highly significant, in the range of 0.10–0.11, confirming that specialization continues to be central. Diversification is consistently negative and significant in this group, with coefficients between -0.24 and -0.11 , suggesting that spreading activity across sectors may dilute resources and hinder the exploitation of scale economies. Structural controls retain their expected positive signs, and the adjusted R^2 of the most complete models reaches roughly 0.22.

In High-medium complexity regions (third quartile), ECI has a positive and significant coefficient only in one specification with full fixed effects (around 0.006), and is small or even negative in the remaining ones. This points to a more ambiguous relationship between complexity and scale at this stage. By contrast, the coefficient on LQ remains very robust, positive and large (around 0.18), indicating that specialization continues to be a strong predictor of productive scale. Diversification has a negative and significant effect in some models, but becomes insignificant once full controls and fixed effects are included. The behaviour of the controls mirrors that of the previous group, and the most complete models

achieve adjusted R^2 values around 0.25.

Table 5 - Results by Level of Economic Complexity of Immediate Brazilian Regions (2006–2023)

	Low	Low-medium	High-medium	High
ECI	0.00651*** (0.00246)	-0.00108 (0.00118)	0.00114* (0.000631)	-0.00116*** (0.000293)
LQ	0.108*** (0.0136)	0.101*** (0.00677)	0.180*** (0.0123)	0.0957*** (0.00818)
Diversification	0.117** (0.0472)	-0.108** (0.0451)	0.0425 (0.0279)	0.161*** (0.0146)
Human Capital	0.00275*** (0.000887)	0.00377*** (0.000344)	0.00349*** (0.000352)	0.00670*** (0.000517)
Trade Openness	0.00262*** (0.000629)	0.00191*** (0.000307)	0.000905** (0.000405)	0.00321*** (0.000558)
Log GDP pc	0.210*** (0.0188)	0.260*** (0.0146)	0.149*** (0.0126)	0.166*** (0.0162)
Log Population	0.135*** (0.0236)	0.205*** (0.0101)	0.0851*** (0.00622)	0.0330*** (0.00521)
Constante	-2.022*** (0.309)	-3.505*** (0.148)	-1.967*** (0.124)	-0.731*** (0.177)
Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
R2	0.138	0.217	0.251	0.256
Adj R2	0.136	0.216	0.250	0.255
Observations	51,344	109,363	128,313	145,671
F-test	61.8	236,3	182.1	194.0

Source: Prepared by the author based on data from RAIS, DataViva, IBGE, and ComexStat.

Note: Standard errors are in parentheses * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Finally, in High complexity regions (fourth quartile), ECI exhibits a striking pattern. In simpler models, the coefficient on ECI is positive and significant, around 0.003 to 0.006, indicating that more complex regions tend to operate with larger plants. Once controls and fixed effects are introduced, however, the coefficient becomes negative and statistically significant (around -0.003 to -0.001). This suggests that among already complex regions, further increases in sophistication are associated with a shift towards knowledge-intensive, technology-based activities that do not require very large establishments to remain efficient. In these contexts, scale is less about physical plant size and more about intangible capital, networks and advanced capabilities. In High complexity regions, LQ remains positive and significant, but the magnitude of its coefficient is smaller than in lower-complexity groups (around 0.09 to 0.10). Diversification has consistently positive and significant coefficients, approximately 0.10 to 0.16, indicating that in these advanced regions a diversified, capability-dense structure supports scale by combining multiple knowledge-intensive activities. Human capital and trade openness continue to show robust positive associations with scale, and the most complete models display adjusted R^2 values in the 0.21 to 0.26 range.

The quartile analysis therefore reveals a clear non-linear relationship between complexity and productive scale. In Low complexity regions, additional complexity and specialization both support scale expansion. In Low-medium and High-medium contexts, the effect of complexity is fragile and depends on complementary structural factors, while diversification can even reduce scale by fragmenting resources. In High complexity regions, the marginal effect of complexity on scale turns negative once structural conditions and fixed effects are controlled for, suggesting that the most sophisticated activities are less dependent on large establishments and more dependent on knowledge intensity, networks and institutional support.

Taken together, the baseline, interaction and quartile results support three main conclusions. First, sectoral specialization remains the primary driver of productive scale in the Brazilian context, in line with the MAR tradition. Second, economic complexity conditions and eventually reshapes this channel, strengthening the effect of specialization in Low complexity settings and attenuating it in High complexity ones. Third, diversification contributes to scale in High complexity regions, where it is

associated with sophisticated activities and dense capability networks, but in intermediate contexts it may reduce the ability to exploit scale economies. These findings underscore the importance of considering the stage of regional sophistication when designing development strategies that seek to combine complex diversification with scale-intensive production.

5. Discussion

The empirical results indicate that the relationship between economic complexity (ECI) and productive scale is neither linear nor spatially homogeneous across Brazilian regions. This heterogeneity resonates with distinct theoretical traditions. On one side, the classical literature on agglomeration and sectoral specialization (Marshall, 1890; Krugman, 1991) emphasizes that the concentration of related activities fosters internal and external economies of scale. On the other, the Jacobian tradition stresses diversified urban structures as engines of innovation, learning and adaptive capacity (Jacobs, 1969).

A central finding is that “sophisticated diversification” does not fully replace specialization as a source of scale economies. While diversified structures may promote experimentation and technological shifts, large-scale production still relies on standardized routines, sunk costs, vertical coordination and sectoral clustering. Diversity supplies dynamism; specialization ensures scalability.

The interaction models reinforce this interpretation. Negative coefficients for the interactions between ECI and both LQ and diversification imply that the scale gains associated with specialization and diversity are attenuated in more complex regions. As regional sophistication increases, productive expansion depends less on exploiting physical scale and more on leveraging accumulated capabilities. Thus, regional development trajectories emerge not from a single mechanism but from changing complementarities between specialization, diversification and complexity.

Further nuances appear when stratifying regions by complexity quartiles. In Low complexity regions, ECI is a robust determinant of productive scale. Even marginal gains in sophistication translate into higher establishment size, consistent with the argument that initial steps in capability accumulation generate strong externalities and encourage plant consolidation (Hidalgo & Hausmann, 2009). Entering more complex activities in shallow capability environments triggers learning, supplier emergence and thicker labor markets, all of which support scale.

In Low-medium complexity regions, the effect of ECI becomes unstable and can even turn negative, whereas LQ remains consistently positive and significant. These territories appear to rely more on sectoral specialization than on complexity for scale gains. Diversification exhibits negative effects, suggesting that spreading limited resources across multiple activities reduces efficiency. This aligns with arguments about the costs of “premature dispersion” when density of related activities is insufficient to sustain multiple trajectories (Hausmann & Klinger, 2006).

In High-medium complexity regions, the link between ECI and scale becomes ambiguous. In some models ECI remains positive, but the effect vanishes once controls are added. This suggests that in partially sophisticated economies, translating complexity into scale requires complementary assets such as skilled labor, integration into external markets, and sufficient market size.

In High complexity regions, the pattern becomes more revealing. Although simpler models suggest a positive effect of ECI, adding controls and fixed effects yields negative and significant coefficients. This supports the idea of diminishing marginal returns of complexity on physical scale at technological frontiers. Knowledge-intensive activities (such as software, IT services or advanced technological niches) operate with lean organizational structures and high human capital intensity rather than large plants (Brynjolfsson & McAfee, 2014). Here, value creation stems from innovation and absorptive capacity (Nelson & Winter, 1982; Cohen & Levinthal, 1990), not from expanding establishment size.

Control variables reinforce these mechanisms. Human capital has a strong positive effect across all specifications, consistent with evidence that skill accumulation fosters productivity, innovation and agglomeration economies (Glaeser & Maré, 2001). Trade openness is also positively associated with scale, confirming the role of external integration in expanding market reach and enabling specialization (Krugman, 1979; 1991). Log GDP per capita and log population show robust positive effects, in line with New Economic Geography arguments linking market size, demand, and urbanization to agglomeration dynamics.

The LQ variable remains consistently positive and significant across all quartiles, confirming specialization as a fundamental driver of scale, in line with MAR-type dynamics. Diversification, in turn, is beneficial only in more complex regions and detrimental in intermediate settings, suggesting that its returns are conditional on capability density. This duality reflects the Jacobian insight that diversity fosters innovation and adaptation, but its contribution to scale depends on structural maturity.

From a policy perspective, the results highlight the importance of sequencing and balance. Less sophisticated regions benefit most from specialization trajectories that consolidate productive routines and reduce costs. As sophistication advances, diversification becomes essential to expand innovation opportunities, reduce vulnerability, and enable technological transitions. An effective regional strategy must therefore combine both dimensions: specialize to scale, diversify to innovate.

Overall, the results show that economic complexity matters for productive scale, but its influence varies with the stage of regional development. In shallow capability environments, rising complexity expands scale. In frontier regions, diversification and innovation become dominant channels, while scale remains anchored in specialized clusters. Rather than weakening the empirical findings, this ambiguity underscores the need for multivariate and stratified approaches to understand regional development processes in Brazil.

6. Conclusion

This article examined how economic complexity, sectoral specialization and productive diversification relate to establishment scale across Brazilian Immediate Regions between 2006 and 2023. The research question asked whether regional complexity supports scale expansion and whether specialization remains the primary channel for scale economies in the Brazilian context. We hypothesized that (i) regional complexity would exert a positive influence on productive scale, particularly in less developed settings, and (ii) specialization would remain the dominant mechanism for scale formation, potentially modulated by sophisticated diversification.

The empirical results partially confirm these hypotheses. In the aggregate, higher economic complexity is associated with larger establishment size, consistent with the view that sophisticated capability structures facilitate productivity and coordination gains. However, the quartile analysis reveals strong heterogeneity: in low-complexity regions, marginal increases in complexity translate into substantial increases in scale, whereas in high-complexity regions the marginal effect turns negative once structural controls are included. This pattern suggests diminishing returns of complexity on physical scale in technology-intensive environments, where efficiency relies less on large plants and more on knowledge-intensive production.

Sectoral specialization, captured by the location quotient, displays positive and highly significant coefficients across all specifications and complexity groups, confirming its centrality as a driver of productive scale. Diversification exhibits nonlinear behavior: negative in low-medium complexity contexts, possibly due to resource dispersion, and positive in high-complexity settings, where sectoral variety reflects complementarities and capability recombination. Interaction models further show that complexity attenuates the scale-enhancing effects of both specialization and diversification, indicating that advanced regions rely less on these individual mechanisms and more on broader capability systems.

Beyond answering the research question, the article contributes in three ways. First, it integrates the literatures on agglomeration, economic complexity and scale economies within a unified empirical framework. Second, it provides novel evidence on how complexity conditions the relationship between specialization, diversification and scale at different stages of regional development. Third, it offers a sub-national application of complexity metrics based on formal employment data, expanding empirical applications to developing-country contexts.

Some limitations should be acknowledged. The RAIS dataset covers only formal employment, excluding informal activities that remain relevant in parts of Brazil. The complexity metric is based on employment data rather than trade or technological indicators, which may under-represent innovation-intensive sectors. The scale measure relies on average establishment size and does not incorporate alternative dimensions such as revenue, cost structures or production volumes. Endogeneity issues, such as reverse causality between complexity and scale, cannot be fully ruled out despite fixed effects and

rich controls.

These limitations open promising avenues for future research. One direction is to incorporate alternative data sources, such as fiscal records, production surveys or microdata on firms, to refine scale measurement. Another is to integrate innovation-related indicators (patents, R&D spending, digital adoption) to better capture knowledge-intensive activities. Methodologically, event-study designs or natural experiments could disentangle causal channels, especially around industrial policies, infrastructure investments or commodity shocks. A further extension involves multi-country comparative analyses to investigate whether the non-linear patterns identified here generalize beyond Brazil.

In conclusion, productive scale in Brazil emerges from a layered interaction between specialization, diversification and sophistication. While specialization remains the primary mechanism through which regions exploit scale economies, complexity plays a conditioning role that strengthens scale formation in less developed settings and moderates it in highly sophisticated ones. For regional policy, these results highlight the importance of sequencing: specialize to build scale and diversify to innovate. A balanced strategy of capability accumulation, specialization and complex diversification appears essential for supporting structural transformation and long-term territorial development.

References

- ALENCAR, Júlia F. L.; FREITAS, Elton; ROMERO, João; BRITTO, Gustavo. Complexidade econômica e desenvolvimento: uma análise do caso latino-americano. *Revista da Sociedade Brasileira de Economia Política*, v. 51, p. 147-167, 2018.
- ARROW, Kenneth J. The economic implications of learning by doing. *The Review of Economic Studies*, v. 29, n. 3, p. 155-173, jun. 1962.
- BALLAND, Pierre-Alexandre; BOSCHMA, Ron; CRESPO, Joan; RIGBY, David L. Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, 2018.
- BARRIOS, Salvador; BERTINELLI, Luisito; STROBL, Eric. Geographic concentration and establishment scale: An extension using panel data. *Journal of Regional Science*, v. 46, n. 4, p. 733-746, 2006.
- BEAUDRY, Catherine; SCHIFFAUEROVA, Andrea. Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, v. 38, n. 2, p. 318-337, 2009.
- BEHRENS, Kristian; DURANTON, Gilles; NICOUD, Frédéric Robert. Productive cities: sorting, selection, and agglomeration. *Journal of Political Economy*, v. 122, n. 3, p. 507-553, jun. 2014.
- BRESSAN, Valéria Gama Fully; BRAGA, Marcelo José; BRESSAN, Aureliano Angel. Eficiência e economia de escala em cooperativas de crédito: uma abordagem de fronteira estocástica de custo com dados em painel. *Advances in Scientific and Applied Accounting*, Viçosa, v. 1, n. 1, p. 1-27, 2008.
- BRASIL. Ministério do Desenvolvimento, Indústria, Comércio e Serviços. ComexStat: estatísticas de comércio exterior. Brasília, DF: MDIC. Disponível em: <https://comexstat.mdic.gov.br/pt/geral>. Acesso em: 11 jun. 2025.
- BRASIL. Ministério do Trabalho e Previdência. Relação Anual de Informações Sociais – RAIS. Brasília, DF: Ministério do Trabalho e Previdência. Disponível em: <https://bi.mte.gov.br/bgcaged/login.php>. Acesso em: 09 jun. 2025.
- BRITTO, Gustavo; ROMERO, João; FREITAS, Elton; COELHO, Clara. The Great Divide: Economic Complexity and Development Paths in Brazil and South Korea. p. 1404-1425. In: *Anais do 1º Encontro da Nacional de Economia Industrial e Inovação [Blucher Engineering Proceedings, v.3 n.4]*. São Paulo: Blucher, 2016.
- BRYNJOLFSSON, Erik; MCAFEE, Andrew. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. Nova York: W. W. Norton & Company, 2014.
- COHEN, Wesley M.; LEVINTHAL, Daniel A. Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly*, v. 35, n. 1, p. 128-152, 1990.
- COMBES, Pierre-Philippe. *Economic Structure and Local Growth: France, 1984–1993*. Economic

- Geography, v. 76, n. 4, p. 301-324, 2000.
- COMBES, Pierre-Philippe; DURANTON, Gilles; GOBILLON, Laurent. The identification of agglomeration economies. *Journal of Economic Geography*, v. 11, n. 2, p. 253-266, 2011.
- DATAVIVA. DataViva: dados sobre economia e emprego no Brasil. Belo Horizonte, MG: DataViva. Disponível em: <https://www.dataviva.info/pt/rankings/location/wages-and-employment>. Acesso em: 15 jun. 2025.
- DIXIT, Avinash K.; STIGLITZ, Joseph E. Monopolistic competition and optimum product diversity. *American Economic Review*, v. 67, n. 3, p. 297-308, 1977.
- DURANTON, Gilles; PUGA, Diego. Nursery cities: urban diversity, process innovation, and the life cycle of products. *American Economic Review*, v. 91, n. 5, p. 1454-1477, 2001.
- FREITAS, Elton Eduardo. Economias externas, atributos urbanos e produtividade: evidências a partir do nível salarial industrial das microrregiões brasileiras, 2000-2010. 2012. 118 f. Dissertação (Mestrado em Economia) – Centro de Desenvolvimento e Planejamento Regional, Faculdade de Ciências Econômicas, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.
- FREITAS, Elton Eduardo. Espaço de Atividades Econômicas e Complexidade Estrutural: uma análise das regiões brasileiras. 2019. Tese (Doutorado em Economia) – Universidade Federal de Minas Gerais, Centro de Desenvolvimento e Planejamento Regional, Belo Horizonte, 2019.
- FREITAS, Elton; ROMERO, João; BRITTO, Gustavo; STEIN, Ana Q.; TORRES, Rodrigo. Dataviva: espaço de atividades e indicadores regionais de complexidade econômica. *Textos para Discussão Cedeplar-UFMG*, n. 657. Belo Horizonte: Cedeplar, Universidade Federal de Minas Gerais, 2023.
- FREITAS, Elton; BRITTO, Gustavo; AMARAL, Pedro. Related industries, economic complexity, and regional diversification: An application for Brazilian microregions. *Papers in Regional Science*, v. 103, p. 100011, 2024.
- FUJITA, Masahisa; KRUGMAN, Paul; VENABLES, Anthony. The spatial economy: cities, regions and international trade. Cambridge: MIT Press, 2001.
- FUJITA, MASAHISA; THISSE, JACQUES-FRANÇOIS. Economics of agglomeration: Cities, industrial location, and globalization. Cambridge University Press, 2011.
- GALA, Paulo; CAMARGO, Jhean; FREITAS, Elton. The Economic Commission for Latin America and the Caribbean (ECLAC) was right: scale-free complex networks and core-periphery patterns in world trade. *Cambridge Journal of Economics*, v. 41, n. 6, p. 1-19, 2017.
- GALA, Paulo; ROCHA, Igor; MAGACHO, Guilherme. The structuralist revenge: economic complexity as an important dimension to evaluate growth and development. *Brazilian Journal of Political Economy*, v. 38, n. 2, p. 219-236, abr./jun. 2018.
- GLAESER, Edward L.; KALLAL, Hedi D.; SCHEINKMAN, José A.; SHLEIFER, Andrei. Growth in cities. *Journal of Political Economy*, v. 100, n. 6, p. 1126-1152, 1992.
- GLAESER, Edward L.; MARÉ, David C. Cities and skills. *Journal of Labor Economics*, v. 19, n. 2, p. 316-342, 2001.
- HARTMANN, Dominik; GUEVARA, Miguel R.; JARA-FIGUEROA, Cristian; ARISTARÁN, Manuel; HIDALGO, César A. Linking economic complexity, institutions, and income inequality. *World Development*, v. 93, p. 75-93, 2016.
- HAUSMANN, Ricardo; KLINGER, Bailey. Structural transformation and patterns of comparative advantage in the product space. Center for International Development Discussion Paper, n. 128, Harvard University, 2006.
- HENDERSON, J. Vernon. Marshall's scale economies. *Journal of Urban Economics*, v. 53, n. 1, p. 1-28, 2003.
- HIDALGO, César A. Economic complexity theory and applications. *Nature Reviews Physics*, v. 3, p. 92-113, 2021.
- HIDALGO, César A.; HAUSMANN, Ricardo. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, v. 106, n. 26, p. 10570-10575, 2009.
- HIDALGO, César A.; KLINGER, Bailey; BARABÁSI, Albert-László; HAUSMANN, Ricardo. The product space conditions the development of nations. *Science*, v. 317, n. 5837, p. 482-487, 2007.
- HOLMES, Thomas J.; STEVENS, John J. Geographic concentration and establishment scale. *Review of Economics and Statistics*, v. 84, n. 4, p. 682-690, 2002.

- HOLMES, Thomas J.; STEVENS, John J. An alternative theory of the plant size distribution with an application to trade. *Journal of Political Economy*, v. 122, n. 2, p. 369-421, 2014.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Divisão regional do Brasil em regiões geográficas imediatas e regiões geográficas intermediárias: 2017. Rio de Janeiro: IBGE, 2017, pág. 20. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?id=2100600&view=detalhes>. Acesso em: 04 de jun. 2025.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Contas Nacionais: Produto Interno Bruto – PIB. Rio de Janeiro, RJ: IBGE. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9088-produto-interno-bruto-dos-municipios.html>. Acesso em: 09 jun. 2025.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Estimativas da População. Rio de Janeiro, RJ: IBGE. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html>. Acesso em: 09 jun. 2025.
- JACOBS, Jane. *The economy of cities*. New York: Vintage, 1969.
- KALDOR, Nicholas. *Causes of the Slow Rate of Economic Growth of the United Kingdom*. Cambridge: Cambridge University Press, 1966.
- KRUGMAN, Paul. Increasing returns and economic geography. *Journal of Political Economy*, v. 99, n. 3, p. 483-499, jun. 1991.
- KRUGMAN, Paul R. Increasing returns, monopolistic competition, and international trade. *Journal of International Economics*, v. 9, n. 4, p. 469-479, 1979.
- LAFOURCADE, Marc; MION, Giordano. Concentration, agglomeration and the size of plants. *Regional Science and Urban Economics*, v. 37, n. 1, p. 46-68, 2007.
- MARSHALL, Alfred. *Principles of economics*. London: Macmillan, 1890.
- MAZZUCATO, Mariana. *The Entrepreneurial State: Debunking Public vs Private Sector Myths*. Londres: Anthem Press, 2013.
- MEALY, Paul; FARMER, J. Doyne; TEYTELBOYM, Alexander. Interpreting economic complexity. *Science Advances*, v. 5, n. 1, eaau1705, 2019.
- MEALY, Paul; FARMER, J. Doyne; TEYTELBOYM, Alexander. What you export matters for emissions: evidence from firm-level data. *Research Policy*, v. 51, n. 2, p. 104384, 2022.
- NEFFKE, Frank; HENNING, Martin; BOSCHMA, Ron. 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, Richard R.; WINTER, Sidney G. *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press, 1982.
- NUNES, Rodolfo Francisco Soares. A complexidade econômica como metodologia de mensuração do desenvolvimento econômico: o caso brasileiro de 2001 a 2016. *Revista Foco*, v. 17, n. 1, p. 1-16, 2024.
- PASCHOALINO, Pietro André Telatin; PARRE, José Luiz; CHAGAS, André Luis Squarize. Diversification, mid-term rainfall and agricultural production. An analysis of the Brazilian Northeast region. *Revista de Economia e Sociologia Rural*, v. 63, e289389, 2025.
- PEREZ, Carlota. *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*. Edward Elgar Publishing, 2002.
- POFFO, Roberto K.; HARTMANN, Dominik; MARIN, Solange Regina. O pensamento de Celso Furtado sobre desenvolvimento econômico e a Abordagem da Complexidade. *SciELO Preprints*, 2023.
- QUEIROZ, Arthur Ribeiro; ROMERO, João; FREITAS, Elton. Relatedness and regional economic complexity: good news for some, bad news for others. *Relatedness and regional economic complexity*, v. 25, p. 1-25, 2024.
- QUINZANI, Marcia Angela Dahmer; BORGES, Fábio. Desenvolvimento e Complexidade Econômica: Uma Análise Comparativa Brasil-China. *Revista Orbis Latina*, p. 151-171, 2020.
- ROMER, Paul M. Increasing returns and long-run growth. *Journal of Political Economy*, v. 94, n. 5, p. 1002-1037, 1986.

- ROMERO, João; SILVEIRA, Fabrício. Mudança estrutural e complexidade econômica: identificando setores promissores para o desenvolvimento dos estados brasileiros. In: LEITE, Marcos Vinicius Chiliatto (org.). Alternativas para o desenvolvimento brasileiro: novos horizontes para a mudança estrutural com igualdade. Brasília: CEPAL, 2019. p. 235-255.
- ROMERO, João; GRAMKOW, Camila. Economic complexity and greenhouse gas emissions. *World Development*, v. 139, 105317, 2021.
- ROSENTHAL, Stuart S.; STRANGE, William C. Evidence on the nature and sources of agglomeration economies. In: HENDERSON, J. Vernon; THISSE, J. F. (org.). *Handbook of regional and urban economics*. Amsterdam: Elsevier, 2004. v. 4, p. 2119-2171.
- SHAPIRO, Carl; VARIAN, Hal R. *Information Rules: A Strategic Guide to the Network Economy*. Boston, MA: Harvard Business School Press, 1999.
- SMITH, Adam. *A riqueza das nações: investigação sobre sua natureza e suas causas*. 1. ed. Londres: W. Strahan e T. Cadell, 1776.
- SOARES, Francisco de Assis; SANTOS, Sandra Maria dos; FREITAS, Elton Eduardo. Especialização industrial e economia de escala: evidências para as microrregiões brasileiras. *Revista Econômica do Nordeste*, v. 39, n. 4, p. 693-713, 2008.
- TACCHELLA, Andrea; CRISTELLI, Matthieu; CALDARELLI, Guido; GABRIELLI, Andrea; PIETRONERO, Luciano. A new metrics for countries' fitness and products' complexity. *Scientific Reports*, [s. l.], v. 2, n. 723, 2012.
- TINOCO, Alexandre. Das economias de aglomeração às externalidades dinâmicas de conhecimento: por uma releitura de São Paulo. *Revista Brasileira de Estudos Urbanos e Regionais*, v. 5, n. 1, p. 47-61, 2003.
- WOOLDRIDGE, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press, 2010.