DOI: 10.5433.2317-627X.2024.v12.n3. 49941

Economia & Região
E-ISSN: 2317-627X

The MAUP and the effect of industrial diversity on regional economic stability

Vinicius Hiago e Silva Gerônimo [1]
https://orcid.org/0009-0007-0331-2324

Marcelo de Oliveira Torres[2]
https://orcid.org/0000-0003-4277-8497

Recebido em: 28/02/2024
Aprovado em:15/04/2024

Abstract

Theoretically, the correlation between industrial diversity and regional economic stability is expected to be positive, suggesting that a region with greater industrial diversification will be less affected by exogenous shocks, resulting in less economic instability. However, historically, empirical studies have produced mixed results, often finding a lack of significance in the correlation. Among the possible reasons for such divergence, this study focuses on the Modifiable Areal Unit Problem (MAUP) and aims to explore the relationship between diversity and stability at different geographical levels, as well as to highlight the possible presence of the MAUP in the data. This study uses spatial econometrics and data from RAIS and the Demographic Census to collect information on Brazilian municipalities, micro-regions and meso-regions between 2010 and 2019. The results show spatial influence at all geographical levels, as well as variation in the magnitude, significance and direction of the correlation depending on the scale used, confirming the study's MAUP hypothesis. In addition, the municipal level was the only one that showed results more consistent with theory in the Brazilian case.

***Keywords**: industrial diversity; economic stability; MAUP; geographic scales; spatial econometrics.*

JEL Codes: C21, O54, R12

[1] Graduação em Ciências Econômicas pela Universidade de Brasília (UnB), Brasília. E-mail: https://orcid.org/0009-0007-0331-2324

[2] Doutorado em Economia Agrícola e dos Recursos Naturais pela Universidade da Califórnia - Davis. Professor da Universidade de Brasília (UnB), Brasília.. E-mail: moliveiratorres70@gmail.com

A MAUP e os efeitos da diversidade industrial na estabilidade econômica regional brasileira

Resumo

A correlação entre diversidade industrial e estabilidade econômica regional é defendida teoricamente como sendo positiva, indicando que uma região que possui uma maior diversificação industrial sofre uma menor interferência de choques exógenos, resultando em uma menor instabilidade econômica. Entretanto, historicamente trabalhos empíricos encontram resultados divergentes, muitas vezes não sendo observada uma significância na correlação. Dentre os possíveis motivos que podem levar a tal divergência, há o Problema de Unidade de Área Modificável (MAUP), foco deste trabalho. O modelo de econometria espacial empregado foi estimado com dados da RAIS e do Censo Demográfico em nível de municípios, microrregiões e mesorregiões brasileiras para os anos de entre 2010 e 2019. Os resultados indicam uma influência espacial em todos os níveis geográficos e apontam para uma variação na magnitude, significância e no sinal da correlação a depender da escala de agregação espacial utilizada, confirmando a hipótese do MAUP no estudo. Além disso, o nível municipal foi o único que apresentou resultados mais coerentes com a teoria no caso brasileiro..

Keywords: industrial diversity; economic stability; MAUP; geographic scales; spatial

Palavras-chaves: diversidade industrial; estabilidade econômica; MAUP; escalas geográficas; econometria espacial.

JEL Codes: C21, O54, R12

Introduction

The sensitivity of statistical analyses to alternative sizes and shapes of geographic units has been discussed for decades, particularly in the field of geography (Dark, 2007). If results change robustly depending on the configuration of the units, conclusions about spatial patterns and causality may become spurious, leading to the Modifiable Areal Unit Problem, better known as MAUP (Openshaw, 1984). This paper investigates whether the MAUP is relevant to the analysis of the relationship between industrial diversity and regional economic stability in the Brazilian context. In doing so it highlights the role of spatial resolution in the estimation of causal effects and in the design of public policies aimed at promoting regional economic stability through diversification of economic activity.

Theoretical models of regional economics show that regions with more stable economic performance tend to have greater industrial diversity (e.g., Chinitz, 1961; Conroy, 1974; Kort, 1981). Unlike specialized regions, diverse regions are better equipped to cope with seasonal cycles in economic activity and the effects of external shocks by smoothing out fluctuations in output and income. Several empirical studies that have attempted to confirm the hypotheses derived from these models, as will be discussed later, typically use only one geographic scale or shape, ignoring the possibility of different spatial configurations. Given that each configuration may provide a different set of information, the estimation of the correlation between industrial diversity and economic stability may be distorted depending on the type of aggregation used.

Following the approach proposed by Anselin (1988) and Chen (2018), this paper incorporate three spatial scales of data aggregation (municipalities, micro-regions, and meso-regions) in the analysis of the impact of industrial diversity on the regional economic performance. The purpose of adopting more than one geographical scale is to check whether the relationship between economic stability and industrial diversity can change in magnitude, sign or significance, indicating the presence of the MAUP.

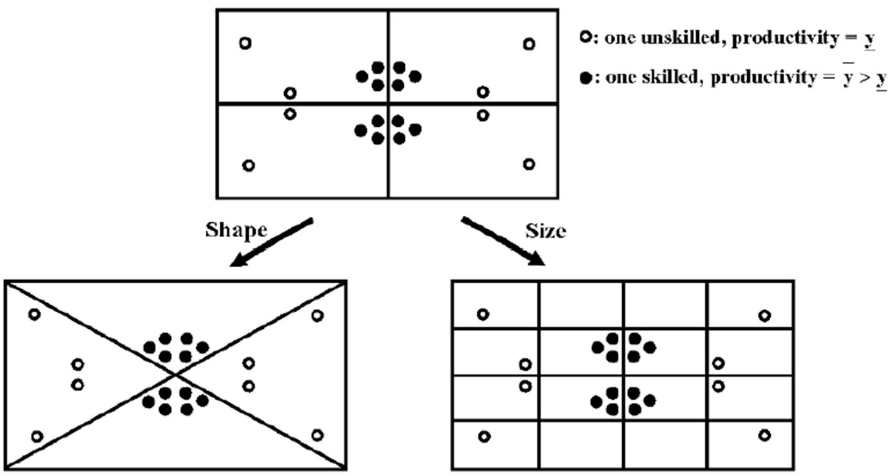
More specifically, data from 5,564 municipalities in Brazil were gathered from secondary sources and subsequently aggregated into different spatial scales, leading up to 558 micro-regions and 137 meso-regions. The contribution of diversity to regional stability is then analyzed with the application of spatial econometric methods at each geographic scale. Information on industrial diversity and control variables are from the year 2010, while data used to feed the economic stability indicator are for the period between 2010 and 2019.

In addition to this introduction, this article is followed by Section 2, which discusses the different results found in key empirical studies on the relationship between diversity and stability, and the Modifiable Areal Unit Problem (MAUP). The third section presents the methodology and the database, along with a spatial an exploratory analysis on the correlation between diversity and stability in Brazil. Section 4 examines the results of the regressions, and the fifth presents the conclusions.

The MAUP, industrial diversity and regional economic stability

The Modifiable Areal Unit Problem (MAUP) is one of the most studied problems in geography and refers to the inconsistency of results when using different spatial aggregations under the same data set. According to Openshaw (1984), the origin of this problem lies on the need to group individual data (such as persons, households, firms, municipalities, etc.) into arbitrary and modifiable zones or regions (such as districts, counties etc.) to facilitate the implementation of statistical research and avoid the need to use microdata.

Figure 1 - The shape and scale problems



Source: Briant *et al.* (2008).

By way of illustration, Figure 1 presents a case of the relationship between employment density and labor productivity, as shown in Briant et al. (2008). The black and white dots represent respectively skilled (more productive) and unskilled (less productive) workers. There are three large rectangles, and workers remain in the same position within all of them. Initially, in the top large rectangle, the number of jobs is equally distributed into four smaller and equally sized rectangles, with each containing the five workers: three skilled and more productive, and two unskilled and less productive. So, the correlation between productivity and employment density is zero. Instead, if the shape of the smaller units changes to equally sized triangles as in the bottom left rectangle, there will be spatial concentration. In this case, triangles with higher density of workers concentrate the skilled and more productive ones and vice-versa, leading up to a positive correlation between productivity and employment density. In the bottom right rectangle, the units' shape is the same as in the top rectangle, but smaller in scale. The spatial concentration is also present, and the correlation remains positive, albeit weaker. So, by simply changing the shape or size of the units, conclusions may change leading up to the shape and scale problems of spatial aggregation.

Examples of studies focused on the MAUP include Resende (2011) and Dapena et al. (2018). The former analyses the effect of spatial correlation on economic growth in Brazil, while the latter examines the effect of employment density on the generation of spatial externalities in Spain. Both studies conclude that the results, in terms of magnitude and statistical significance, vary depending on the geographical shape and scale used. In Dapena et al. (2018), the spatial database was sufficiently disaggregated allowing for the use of different geographical delineations and the selection of the most appropriate scale and shape.[3].

In this study, the MAUP is examined in the context of the influence of industrial diversity on regional economic stability. As noted in the previous section, while industrial diversity is theoretically correlated with greater regional economic stability, empirically this correlation is less robust. For example, in testing the hypothesis of a positive correlation between industrial diversity and economic stability in 106 MSAs[4] in the U.S., Kort (1981) found that regions highly specialized in the production of specific industries, such as Kenosha and Gary, had higher instability indices, while the Chicago and New York MSAs were quite diversified and stable. In other words, it was found that regions with diversified economies tended to have higher stability indices.

Similarly, the study by Conroy (1975), which followed an industrial portfolio approach to measure employment variance, found that the variances had a strongly significant relationship with instability indices. Specifically, 42% of regional economic instability was explained by the variance of the industrial portfolio index. This positive correlation was also confirmed in Malizia and Ke (1993) and Trendle and Shorney (2004).

In contrast, Keinath (1985), when examining the influence of industrial diversity on the economic performance of 183 American economic areas, concluded that diversity is not relevant. Attaran (1986) reached a similar conclusion when he analyzed

[3] This approach is "more in line with recent developments in New Economic Geography (NEG), where areas within non-administrative regions such as NUTS-2 or NUTS-3 play a significant role" (Dapena et al., 2018).

[4] An MSA, or Metropolitan Statistical Area, is defined by the United States Office of Management and Budget (OMB) as a "central area containing a substantial population nucleus, together with adjacent communities having a high degree of economic and social integration with that nucleus." According to the standards set by the agency in 2010, an MSA must contain at least one urbanized area with a population of 50,000 or more.

the correlation between industrial diversity and variables such as growth and instability of unemployment levels in the 50 US states between 1972 and 1981. On the other hand, Mizuno (2006), while demonstrating that industrial diversity plays an important role in economic stability, has concluded that factors such as the industrial sector's participation in the economy and the educational level of the workforce are even more important.

More recently, Deller and Watson (2016), using data from U.S. counties from 2005 to 2012 and a geographically weighted regression model, indicate that the significance of the correlation between industrial diversity and economic stability varies depending on the location of the county. That is, although the results support the hypothesis of a positive correlation in western and eastern regions, this relationship was not found in many regions in the central United States.

Several reasons can be cited as potential drivers of these divergences. As pointed out by Siegel (1995), Wagner (2000), Kort (2003), and Conroy (1975), the choice of the diversification indicators, the methodology used and the shape and scale of the decision-making units, all can lead to different results and conclusions. As indicated above this paper focuses on the latter and uses spatial regression methods to account for the potential spatial dependence of regional data according to the works of LeSage (2009) and Deller and Watson (2016).

Methodology

In the context of this study, it is assumed that the correlation between economic stability and industrial diversity in a given region, as observed by Deller and Watson (2016) and Chen (2018), may be influenced by factors belonging to its neighbors that are not randomly distributed in space. For example, empirical studies such as Resende (2011) show that local socioeconomic and infrastructural conditions can systematically imply costs and benefits to neighboring regions. For these reasons, the use of spatial econometric models is preferred over traditional models such as the ones based on linear regressions.

As emphasized by LeSage (1999), these traditional models have limitations in dealing with the inherent spatial dependence and heterogeneity in the data generation process.[5] That is, they fail to account for the spatial dimension of economic, social, and demographic activities and the occurrence of spillovers and interactions between observations[6]. Formally, this means that in models based on linear regressions with ordinary least squares estimators, one can no longer ensure that $E(X) = 0$ and $E\,(\varepsilon\varepsilon') = \sigma^2 I$. That is, strict exogeneity and independence of errors cannot be assumed anymore, leading to inconsistent and inefficient estimators [7].

Following these studies, we build three different spatial econometric models: Spatial Autoregressive (SAR), Spatial Error (SEM), and Spatial Durbin (SDM). Tests are then performed in each geographical unit studied, looking for the one with the best explanatory power. SAR incorporates the spatial interaction between the dependent variable of one region and the dependent variable of neighboring regions, formalizing

[5] As will be seen later through the exploratory analysis carried out in Section 3.2.1, locational factors are indeed relevant for the analysis of patterns of economic stability and industrial diversity of geographical units, suggesting the need for spatial econometric estimation methods.

[6] In non-spatial data, observations from different individuals are considered independent of each other. However, as Tobler (1970) points out in his 1st Law of Geography: "*everything is related to everything else, but near things are more related than distant things*". That is, spatially distributed phenomena are not independent.

[7] For a formal derivation of this result, see Elhorst (2014).

the notion of spillovers (Anselin, 2002). In the context of this study, the SAR model can be described by the following equation:

$$REI = \rho WREI + \beta_{hhi}HHI + X\beta + \varepsilon. \quad (4)$$

Where *REI* is an indicator of economic instability, *HHI* is an indicator of industrial diversity, with an associated parameter β_{hhi}. **X** is a vector of control variables, and β is a vector of parameters. ρ is the autoregressive parameter, and W is the spatial weight matrix[8]. Thus, SAR controls for strict exogeneity as it allows for the regional level of economic instability to be explained by the economic instability of neighboring regions.

SEM assumes that observations are interdependent due to unobserved variables. Therefore, unobserved characteristics that may affect neighboring regions are controlled by considering the interdependence of the errors in the model. Formally, it can be expressed by the following equation:

$$REI = \beta_{hhi}HHI + X\beta + u,$$
$$u = \lambda Wu + \varepsilon. \quad (5)$$

Where the variables *REI* and *HHI* are defined as above, β_{hhi} is a parameter, **X** is a vector of control variables, and β is a vector of parameters. In the error term, W is the spatial weight matrix, λ is the autoregressive parameter, and ε is the random error.

Finally, SDM attempts to capture spillover effects from both the explanatory variables and the dependent variable. It allows for the inclusion of direct and indirect spatial effects using spatial lags of the dependent and independent variables. These spatial lags allow for the influence of omitted spatial variables that may be correlated with both the dependent and explanatory variables. Its formal representation is described by the following equation,

$$REI = \rho WREI + \beta_{hhi}HHI + X\beta + \varphi_{hhi}WHHI + WX\theta + \varepsilon. \quad (6)$$

Where ρ is the autoregressive parameter and W is the neighborhood matrix. *REI*, *HHI*, β_{hhi}, **X** and β are still defined as above. θ is a vector of parameters measuring the marginal effect of the neighbors' explanatory variables and φ is a parameter measuring the marginal effect of the neighbors' *HHI*.

In all models, the neighborhood matrix identifies the observations that should be considered as neighbors of a given region. *W* is a positive *n* x *n* matrix, in which each row *i* contains a non-zero element w_{ij} that defines *j* as *i*'s neighbor. Thus, the element *j* is assigned a value 1 if it is identified as a neighbor of *i*, otherwise 0. Neighbors can be defined based on contiguity if the region is adjacent to another region or has one of the polygon's vertices in common. Alternatively, a region can be considered a neighbor of another region if it lies within a certain area whose extent is specified by the researcher.

In this paper, the *k* nearest neighbor matrix was used. That is, the *k* regions with the centroid closest to the centroid of its polygon are the neighbors of a region *i*. The neighborhood matrix can then be understood as a weight matrix that forces

[8] Both indicators, control variables, and the neighborhood matrix are explained in more detail at the end of this section.

observations to have the same number of geographic units that affect their economic instability and avoids the so-called islands, since all observations will have *k* neighbors, regardless of distance and contingency (Anselin, 2002; Chen, 2018; Resende, 2011; Neves et al., 2019; Souza et al., 2021). In the baseline estimates, the 10 nearest neighbors were considered, but simulation results using matrices with 5 and 15 nearest neighbors are also presented and discussed.

The indicators for industrial diversification and economic instability

The explanatory variable *HHI* in the spatial models represented by equations (4), (5) and (6) is the *Herfindahl-Hirschman* index applied to measure the industrial diversity of geographical units. More specifically,

$$HHI = \sum_{j=1}^{N} (\frac{o_{ij}}{o_i})^2. \qquad (8)$$

Where *N* is the number of economic sectors in region *i*, o_{ij} is the number of employees in sector *j* of region *i* and o_i is the total number of employees in region *i* (Wagner, 2000; Deller and Watson, 2016; Chen, 2018).

According to Eq. (8), the closer the *HHI* value is to 1, the greater the economic concentration of the region, i.e. the greater the number of jobs concentrated in a few sectors in relation to the number of sectors present in the geographical unit. Consequently, the lower the value of the index, the better the distribution of jobs in the region are, i.e. the more diversified is its economic activity.

For the regional economic instability indicator, *REI*, we follow the version proposed by Malizia and Ke (1993), where the average of the total deviation of employment from its linear trend is divided by the linear trend. That is,

$$REI = \left\{ \sum_{i=1}^{N} [\frac{(o_{it} - o_{it}^T)}{o_{it}^T}]^2 / T \right\}^{1/2}. \qquad (9)$$

Where O_{it} represents the total number of employed persons in region *i* in period *t*, O_{it}^T represents the linear trend of the total number of employed people in region *i* in period *t* and *T* represents the number of analyzed periods (years of 2010 to 2019). The greater the deviation of the number of employed persons from its trend, the higher the value of the *REI* index, meaning a greater economic instability. Conversely, the lower the *REI* value, the more economically stable is the region under study.

Other control variables, database and descriptive statistics

In the estimation of the models presented in (4), (5) and (6), three geographical scales were used, including 5,564 municipalities, 557 micro-regions and 137 meso-regions[9]. The construction of the maps was based on *shapefiles* of Brazilian municipalities, micro-regions and meso-regions available on the IPEAGEO website[10].

[9] According to IBGE, the number of official geographical units until in 2010, was 5,564 municipalities, 558 micro-regions and 137 meso-regions, although some units were excluded from the analysis due to lack of data.
[10] https://www.ipea.gov.br/ipeageo/malhas.html.

All data considered were obtained at the municipal level and then scaled up to micro-regions and meso-regions.

The control variables in the vector **X** in (4), (5) and (6) described in Table 1 above were obtained from the "*Atlas do Desenvolvimento Humano no Brasil*", which consolidates data on health, education, income, work, housing, political participation, and the environment from 5,570[11] municipalities, based on the 2010 IBGE[12] Census. It is worth highlighting that the choice of variables in **X** was made based on previous works on the topic (Chen, 2018; Watson e Deller, 2017; Resende, 2011; Deller e Watson, 2016), after accounting for factors related to their statistical significance and data availability.

Table 1 – Description of the Variables

Variable	Description	Source
REI	Average deviation of total employment from its linear trend divided by the linear trend, 2010-2019.	RAIS-MTE
HHI	Industrial diversity index using the *Herfindahl-Hirschman* Index, 2010.	RAIS-MTE
POB	Percentage of the population with per capita household income equal to or less than R$140.00 per month, 2010.	Censo-Atlas
SUP25	Percentage of the population aged 25 and older with at least some tertiary education, 2010.	Censo-Atlas
POP18	Percentage of total population 18 years and older, 2010.	Censo-Atlas
TRABSC	Percentage of employed persons aged 18 and older without a formal contract, 2010.	Censo-Atlas
DSPG	Unemployment rate – 18 years and over, 2010.	Censo-Atlas
NO	Dummy variable for regions in the North.	Censo-Atlas
NE	Dummy variable for regions in the Northeast.	Censo-Atlas
CO	Dummy variable for regions in the Centre-West.	Censo-Atlas
SE	Dummy variable for regions in the Southeast.	Censo-Atlas

Source: Own elaboration.

The *HHI* variable is calculated using data on employment from the *Relação Anual de Informações Sociais*[13] (RAIS) from 2010 and data on economic sectors from CNAE[14] division 2.0. The *REI* variable is also calculated based on the RAIS data, considering the years from 2010 to 2019.

[11] Six municipalities were excluded from the sample due to a lack of necessary data to construct REI and HHI.

[12] IBGE, the Brazilian Institute of Geography and Statistics, is a Brazilian government institution responsible for producing and disseminating statistical and geographical information about the national territory, including the collection, organization and analysis of data on demography, economics, the environment and geography.

[13] The RAIS (Annual Social Information Report) is an ancillary obligation of Brazilian employers, required by the Ministry of Labor and Employment, and is an annual statement that contains detailed information on employees and employment relationships, such as remuneration, hours worked, among other relevant data.

[14] The CNAE 2.0 (National Classification of Economic Activities) is a system used to classify the economic activities carried out by companies in Brazil. It is drawn up and maintained by the IBGE (Brazilian Institute of Geography and Statistics) in partnership with other government bodies and private entities. In 2007, version 2.0 of the previous classification (CNAE 1.0) was created, with two objectives in mind: updating the national classification in the light of revision 4 of the International Uniform Industrial Classification - CIIU/ISIC.

Table 2 below presents the descriptive statistics of the variables at the municipal, micro-regional and meso-regional levels.

Table 2 –Descriptive statistics of variables

Variable	Municipalities	Micro-regions	Meso-regions
Observations	5564	558	137
REI			
Max	3.162	0.410	0.249
Min	0.010	0.013	0.019
Mean	0.098	0.049	0.039
S.D	0.118	0.037	0.023
HHI			
Max	1.000	0.917	0.826
Min	0.043	0.044	0.043
Mean	0.379	0.225	0.176
S.D	0.257	0.184	0.153
POB			
Max	0.785	0.653	0.604
Min	0	0.011	0.019
Mean	0.231	0.222	0.203
S.D	0.179	0.167	0.152
SUP25			
Max	0.336	0.239	0.239
Min	0.003	0.020	0.024
Mean	0.055	0.072	0.086
S.D	0.032	0.037	0.039
POP18			
Max	0.834	0.778	0.763
Min	0.278	0.492	0.514
Mean	0.685	0.682	0.683
S.D	0.057	0.053	0.052
TRABSC			
Max	0.622	0.457	0.364
Min	0.030	0.065	0.088
Mean	0.252	0.240	0.227
S.D	0.098	0.076	0.068
DSPG			
Max	0.384	0.202	0.136
Min	0	0.010	0.027
Mean	0.061	0.067	0.069
S.D	0.036	0.026	0.021

Notes: Correlation coefficients are significant at the 1% level. 14 municipalities have an HHI of 1 due to the fact that all of their jobs were allocated to “Public administration, defense and social security”.

It can be observed in Table 2 that the mean values of *HHI* and *REI* decrease as the geographical scale becomes more aggregated. From 0.176, to 0.379 for *HHI* and from 0.039 to 0.098 for *REI*, indicating that smaller units have on average greater instability and less industrial diversity. The same is true for the standard deviation of both variables. The mean values and standard deviations of the control variables POB, SUP25, POP18, TRABSC and DSPG show little variation when the geographic scale either at the municipality, micro-region, or meso-region levels.

Spatial autocorrelation indicators

An important indicator of the relevance of location as an explanatory factor is the existence of spatial autocorrelation, which occurs when similarity of location (neighbors or spatial proximity) is matched by similarity of values (correlation). Generally, the object of an autocorrelation analysis is the dependent variable specified in econometric estimation models, but it can also be applied to any variable distributed across a given geographic area. The presence of spatial autocorrelation suggests non-randomness of the sample and implies the need to account for the spatial dimension, as in the models specified in (4), (5) or (6).

In recent decades, the development of a set of techniques for describing and visualizing the distribution of spatial autocorrelation patterns gave rise to the Exploratory Spatial Data Analysis (ESDA), (Anselin, 1996). Its main tool is the global Moran index, given by:

$$I = \frac{\sum_{i,j}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x}) / \sum_{i,j}^{n} w_{ij}}{\sum_{i}^{n}(x_i - \bar{x})^2 / N}. \tag{10}$$

Where x_i is the value of the variable *x* in municipality *I*, $\bar{x}$ is the average of *x* over the entire study area and w_{ij} is an element of the spatial weight matrix. If municipality *i* shares a common border with municipality *j*, then w_{ij} = 1, otherwise w_{ij} = 0. *I* varies between -1 and 1, where 1 indicates a perfect positive autocorrelation and -1 indicates a perfect negative autocorrelation. The closer to 0, the greater the location independence of the attributes. It is important to emphasize that the global index does not identify local patterns present in the map, it only returns a single index indicating the presence of a more general autocorrelation of the variables.

Table 3 shows the global Moran Indices[15] calculated for the *HHI* and *REI* variables. Both have positive and significant values[16], but while the *REI* Moran index has smaller values and less dispersion, the *HHI* index has greater dispersion and larger values. In any case, the spatial dependence of diversity seems to be greater than that of regional stability at all the scales studied.

Although the above results indicate a spatial dependence in the *HHI* and *REI* variables, they do not allow us to infer where in the Brazilian territory this dependence occurs. To identify local correlation patterns, it is necessary to use the local Moran index, or LISA[17], which can be seen in Figure 2 ~~below~~.

[15] For a more formal derivation of the Global Moran Index, see CHEN, 2013.
[16] Being positive, the Moran Index indicates that, on average, municipalities with high (low) levels of industrial diversity and instability are neighbors to municipalities also with high (low) levels of these same variables.
[17] For a better description of LISA, see Anselin, 1995.

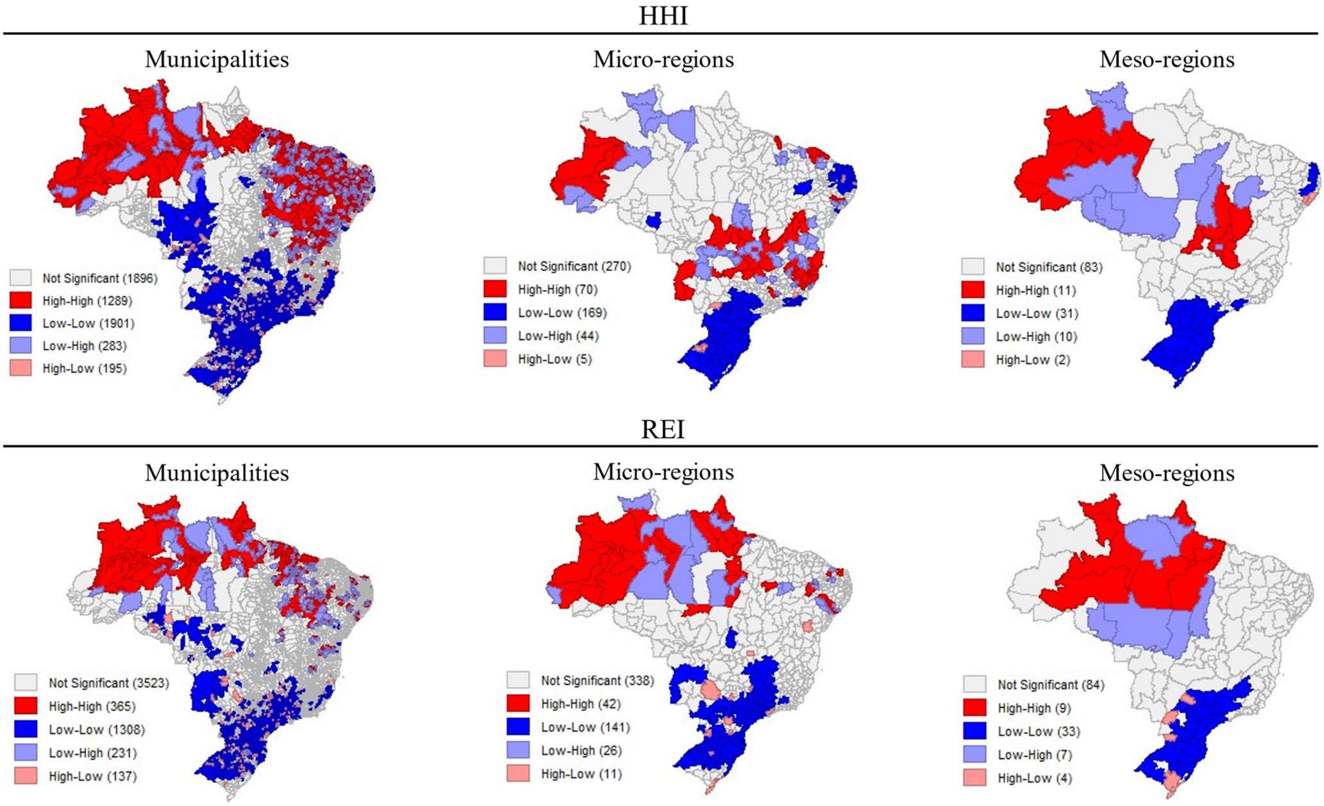
Table 3 – Global Moran indices for *HHI* and *REI*

Scale	Municipalities	Micro-regions	Meso-regions
HHI	0.540	0.289	0.112
REI	0.176	0.196	0.182

Note: All statistics are significant at the 1% level.

Although the above results indicate a spatial dependence in the *HHI* and *REI* variables, they do not allow us to infer where in the Brazilian territory this dependence occurs. To identify local correlation patterns, it is necessary to use the local Moran index, or LISA[18], which can be seen in Figure 2 below.

Figure 2 – LISA Cluster Maps



Source: Own elaboration.

The regions in dark and light blue have respectively *HHI* and *REI* levels below the average. That is, the dark ones have a higher industrial diversity and are close to highly diverse regions, and the light ones have a lower economic instability and are close to regions with lower levels of instability. The regions in dark and light red, on the other hand, have respectively *HHI* and *REI* levels above the average. That is, the dark ones have a lower industrial diversity index and are situated close to regions with a lower diversity index, and the light ones, with higher economic instability, are located near regions with a higher level of instability.

[18] For a better description of LISA, see Anselin, 1995.

The southern and southeastern areas of Brazil have a greater occurrence of *Low-Low* autocorrelation in both the *HHI* and *REI* at the three scales. That is, the values of the instability and diversification index have values below the national average and are around of similar regions. This suggests that between 2010 and 2019 there was less instability and a higher rate of employment diversification between sectors in these locations, which supports the diversification-instability correlation hypothesis.

The Northeast and specially the North regions have an opposite characteristic, i.e. a greater occurrence of the *High-High* autocorrelation pattern in both the *HHI* and *REI* at the three scales. That is, the values of the instability and diversification index have values above the national average and are located around similar regions. These regions are therefore characterized by a greater instability between 2010 and 2019 and a lower rate of employment diversification among the sectors of these economies, which also supports the diversity-instability correlation hypothesis.

These findings are in line with those in Vreyer and Spielvogel (2005). They corroborate the hypothesis of economic polarization in the Brazilian territory and show a stark difference in performance of the regional economies of the Brazilian South *vis-à-vis* the ones in the North/Northeast. The observed clustering configuration suggests the existence of spatial spillover effects in relation to the economic diversity and stability, the magnitude of which being influenced by the regional scale (Resende, 2011). In sum, they reveal the presence of spatial correlation at the three geographic scales and suggest the need to control for spatial factors in the estimation of the industrial diversity effects on economic stability.

Estimates and results

The three main strategies for estimating economic diversity effects on stability are defined in (4), (5) and (6). Tests to specify the strategy that better suit each geographic scale were performed by comparing the statistics produced by the *Likelihood Ratio Test (LR)* applied to the SDM, SEM and SAR models. Specifically, SDM was considered the unrestricted model, SEM and SAR the restricted models. For clarity, the test comparing the SDM and SEM is called *LR Error* and the test comparing SDM and SAR, *LR Lag*.

LR Error and *LR Lag* null hypotheses state that the SDM is statistically equivalent to SEM and SAR, thus making the restricted model preferable. The alternative hypothesis is that SDM is statistically preferable to the others. Therefore, a rejection of the null hypothesis by the *LR Lag* alone would indicate that SEM would be preferred. Now, if the null hypothesis is rejected only due to the *LR Error*, then SAR would be preferred[19]. The test results are shown in Table 4 below and they reveal that *SDM* is preferred in all geographic units, as the test rejects the null hypothesis for both *LR Lag* and *LR Error* at all scales. Conclusions are henceforth based on the results from the estimation of the SDM model, the preferred specification.

[19] For a better description of the Likelihood Ratio Test, see Elhorst, 2014.

Table 4 - Comparison of Likelihood Ratio Tests

Tests	Municipalities	Micro-regions	Meso-regions
LR Lag	32.41***	19.31**	29.81***
LR Error	34.04***	17.65*	18.77**

Notes: Standard errors in parentheses. * significant at 10%; ** significant at 5%; and *** significant at 1%.

Focusing on main variable of interest, *HHI*, Table 5 below shows striking differences in the instability-diversity relationship depending on the scale of data aggregation. At the municipal level, the coefficient of *HHI* is positive and significant, indicating that greater industrial concentration leads to greater instability. The positive and significant value of the coefficient associated with *W*HHI*, which captures the average values of industrial diversity in neighboring municipalities, indicates that economic instability in a given municipality increases when industrial concentration in its neighboring regions increases, a spillover effect. Furthermore, *ρ* is also positive and significant, corroborating the hypothesis that the level of economic instability in the region has a positive correlation with the level of instability in neighboring regions, as suggested by the global Moran's I.

Table 5 - Spatial regressions results – SDM model

Estimation	Municipalities	Micro-regions	Meso-regions
HHI	0.046*** (0.009)	-0.002 (0.008)	-0.026** (0.011)
POB	0.162*** (0.027)	0.086*** (0.031)	0.047 (0.042)
SUP25	-0.216*** (0.056)	0.052 (0.069)	0.018 (0.082)
POP18	-0.088 (0.056)	-0.253*** (0.091)	0.160 (0.109)
TRABSC	-0.002 (0.024)	-0.062 (0.040)	0.072 (0.058)
DSPG	0.101* (0.056)	0.011 (0.080)	-0.063 (0.105)
NO	-0.006 (0.043)	-0.021 (0.025)	0.027** (0.020)
NE	0.022 (0.039)	0.008 (0.024)	-0.007 (0.022)
CO	0.012 (0.035)	-0.006 (0.021)	0.014 (0.018)
SE	0.026 (0.029)	-0.009 (0.019)	0.005 (0.016)
W*HHI	0.074*** (0.021)	-0.022 (0.022)	-0.123*** (0.046)
W*POB	-0.148*** (0.051)	0.048 (0.079)	-0.129 (0.213)
W*SUP25	0.117 (0.143)	-0.061 (0.181)	-0.350 (0.409)

W*POP18	-0.200** (0.099)	0.419** (0.175)	0.277 (0.399)
W*TRABSC	-0.122*** (0.039)	0.062 (0.077)	0.371** (0.171)
W*DSPG	0.046 (0.099)	0.169 (0.163)	-0.510 (0.325)
Constant	0.245*** (0.065)	-0.103 (0.115)	-0.282 (0.309)
Ρ	0.236*** (0.027)	0.096 (0.115)	-1.112*** (0.259)
Observations	5564	558	137

Notes: Standard errors in parentheses. * significant at 10%; ** significant at 5%; and *** significant at 1%.

At the micro-regional level, the effect of *HHI* on *REI* is not significant, and at the meso-regional level, negative and significant at 1%. In other words, greater industrial diversity increases the economic instability of the region, which is the opposite of what is expected from the theory. The coefficient of *W*HHI* is also negative and significant, indicating that the greater industrial diversity of the region reinforces the economic instability of the neighbors. Another conflicting result regarding the two geographical scales refers to ρ, which was negative and significant at a 1% level. It suggests that the economic instability in a given region decreases when its neighbors become more economically instable.

Marginal, direct and indirect effects of industrial diversity on regional economic instability

As noted by LeSage and Pace (2009), the coefficients of models with lagged variables, such as the SDM, should not be interpreted as marginal effects.[20] Marginal effects can be decomposed into direct, indirect and total. The direct marginal effect is the effect of a variation in the diversity of region *i* on the economic instability of region *i*. The indirect effect, in turn, reflects the effect of the variation in diversity in region *i* on the economic instability of all the other regions, whether neighboring or not. The total effect represents the sum of the direct and indirect effect. The results are shown in Table 6 below.

Table 6 - Estimation of direct, indirect and total *HHI* effects – SDM model

Estimation	Municipalities	Micro-regions	Meso-regions
Direct effect	0.047***	-0.002	-0.02*
Indirect effect	0.109***	-0.024	-0.05**
Total effect	0.157***	-0.026	-0.071***

Notes: * significant at 10%; ** significant at 5%; and *** significant at 1%.

[20] For a formal derivation of the calculation of marginal effects, see Elhorst (2014).

Regarding municipalities, the direct effect of *HHI* was significant with a value of 0.047, while the indirect effect was significant with a value of 0.109. This means that an increase of one unit in the *HHI* in given region increases its own *REI* by 0.047, on average, everything else constant. Furthermore, these increase in the *HHI* increases the *HHI* in other regions altogether by 0.109, resulting in a total effect on the *REI* of 0.157. In other words, an increase in industrial diversity in one region affects the economic stability of that region as well as of other regions since the industrial diversity of neighbors also influences their stability (as can be inferred from the SDM model). In other words, an increase in industrial concentration by one unit of measurement increases the economic instability of regions by 0.157. With respect to the other levels of aggregation, the total effect can be decomposed in the same way into direct and indirect effects, all of which are negative and significant only when at the meso-regional level. As discussed above, these results contradict what is predicted by theory by suggesting that greater industrial concentration reduces economic instability.

It is important to note that the indirect effect already internalizes the spillover effects on other neighbors and the direct effect already accounts for the feedback effect. That is, the direct effect captures both the initial impact and the impact that the *HHI* caused in the other regions and subsequently returned to the region of origin (LeSage and Pace, 2009). It is also worth mentioning that the larger magnitude observed in the indirect effects is consistent with Chen (2018), who reported the same phenomenon in most of his analyses.

Sensitivity analysis

To perform a sensitivity analysis of the results with respect to the definition of neighborhood, both the spatial autocorrelation patterns and the estimates were re-evaluated with matrices of 5 and 15 nearest neighbors. Table 7 shows that the global Moran Indices for *HHI* and *REI* are positive and similar in the three geographic units, regardless of the number of neighbors in **W**. It is also possible to note that, in general, the greater the number of regions considered as neighbors in the matrix, the lower the index value.

Table 7 - Global Moran index for different neighborhood matrices

Variable	Matrix	Municipalitie s	Micro-regions	Meso-regions
HHI	5 nearest	0.558	0.346	0.229
	10 nearest	0.540	0.289	0.182
	15 nearest	0.530	0.263	0.172
REI	5 nearest	0.169	0.229	0.123
	10 nearest	0.176	0.196	0.112
	15 nearest	0.172	0.179	0.108

Notes: All statistics are significant at the 1% level.

For the marginal effects of *HHI* on *REI*, Table 8 shows that, when estimated at the municipal level, the direct effect remains relatively stable, but the indirect effect increases with the increase in the number of regions considered as neighbors. At the micro-regional level, the effects are negative and insignificant in all the matrices, and at the meso-regional level, they are either negative and significant with the 10 nearest matrix or positive and significant, for the direct effect, with the 15 nearest matrix.[21]

Table 8 - Marginal *HHI* effects under different neighborhood matrices – SDM model

Matrix	Estimation	Municipalities	Micro-regions	Meso-regions
5 nearest				
	Direct effect	0.049***	-0.001	-0.010
	Indirect effect	0.086***	-0.015	-0.035
	Total effect	0.136***	-0.016	-0.046
10 nearest				
	Direct effect	0.047***	-0.002	-0.02*
	Indirect effect	0.109***	-0.024	-0.05**
	Total effect	0.157***	-0.026	-0.071***
15 nearest				
	Direct effect	0.047***	-0.003	-0.007
	Indirect effect	0.131***	-0.032	0.057*
	Total effect	0.178***	-0.036	0.049

Notes: * significant at 10%; ** significant at 5%; and *** significant at 1%.

This sensitivity analysis is consistent with the results of the previous section by highlighting the MAUP in the data, as the sign, magnitude, and significance of the *HHI* coefficient, and of other coefficients as well, change considerably depending on the type of aggregation used. These results contrast to Chen (2018)'s that finds MAUP evidence only in relation to statistical significance and magnitude, not the sign, which was always positive at all levels of data aggregation. In this study, the positive and more stable correlation defended by the theory was observed only at the municipal level.

Conclusions

This paper aimed to study the correlation between industrial diversity and regional economic stability at different geographic scales in Brazil and test for the presence of the Modifiable Area Unit Problem, better known as MAUP. The fact that the magnitude, significance and sign of the correlation change considerably given the level of data aggregation indicates that the spatial pattern of the data interferes with the results of the study and emphasizes the importance of considering alternative geographic scales in regional studies.

[21] The significance, magnitude and sign for other matrices can be provided upon request. In general, for matrices of size between 5 and 10 and between 10 and 15, the direct effects remain stable and significant in the municipal case, and the indirect effects increase with the number of neighbors. At the micro-regional level, all effects remain negative and insignificant. At the meso-regional level, the effects are generally negative and insignificant, but both the signs and the significance can vary depending on the number of neighbors considered.

In particular we find that the pattern of distribution of regional instability across space is less dependent on location at all geographic scales in contrast to industrial diversification which presents a much higher spatial autocorrelation and a higher clustering. The exploratory spatial analysis of the data confirms the conventional wisdom that areas in the southeast of Brazil or further south are more industrially diverse and clustered with similar neighboring regions. In addition, these regions also showed less economic instability between 2010 and 2019, confirming the diversity-instability correlation hypothesis. The Northeast region, and specially the North region, have the opposite characteristic, that is, they are in general less industrially diverse and clustered with similar neighbors. They also contain more areas with high economic instability that are located near other areas also more economically unstable.

The econometric results at the municipal level reveal positive impacts of industrial diversity on the economic stability with significant feedback and spillovers effects. At other aggregation scales, the model estimation has yielded confounding results. These findings indicate the presence of the MAUP, which was reinforced in the sensitivity analysis performed with alternative neighborhood matrices. For future work, it is suggested to include more than one diversity index and to include spatial clusters in the analysis.

The above results highlight the need for public policies aimed at covering a wider territorial scope, especially in the North and Northeast, given the spillover effects and spatial autocorrelation. In other words, isolated municipal policies aimed at promoting industrial diversity or regional economic stability in a single municipality may not be effective, since the performance of regions in terms of these variables is partially dependent on the performance of neighboring municipalities. Since these two Brazilian regions are mostly characterized by more unstable and poorly diversified agglomerations, their municipalities may find themselves in a limited and persistent development cycle.

References

ANSELIN, Luc. **Spatial econometrics: methods and models**. Springer Science & Business Media, 1988.

ANSELIN, Luc. Interactive techniques and exploratory spatial data analysis. 1996.

ANSELIN, Luc. Under the hood issues in the specification and interpretation of spatial regression models. **Agricultural economics**, v. 27, n. 3, p. 247-267, 2002.

ANSELIN, Luc. Spatial externalities, spatial multipliers, and spatial econometrics. **International regional science review**, v. 26, n. 2, p. 153-166, 2003.

ATTARAN, Mohsen. Industrial diversity and economic performance in US areas**. The Annals of Regional Science**, v. 20, n. 2, p. 44-54, 1986.

CARVALHO, A.; DA MATA, Daniel; RESENDE, Guilherme Mendes. Clusterização dos municípios brasileiros. **Dinâmica dos municípios**, p. 326-368, 2007.

CHEN, Jing. Economic Diversity and Regional Development: Geographical Scale, Structural Measurement, and Modeling Method. **West Virginia University**, 2018.

CHEN, Yanguang. New approaches for calculating Moran's index of spatial autocorrelation. **PloS one**, v. 8, n. 7, p. e68336, 2013.

CHINITZ, Benjamin. Contrasts in agglomeration: New york and pittsburgh. **The American Economic Review,** v. 51, n. 2, p. 279-289, 1961.

CONROY, Michael E. The concept and measurement of regional industrial diversification. **Southern Economic Journal**, p. 492-505, 1975.

DARK, Shawna J; BRAM, Danielle. DARK, Shawna J.; BRAM, Danielle. The modifiable areal unit problem (MAUP) in physical geography. **Progress in physical geography**, v. 31, n. 5, p. 471-479, 2007.

DÍAZ DAPENA, Alberto; FERNÁNDEZ VÁZQUEZ, Esteban; RUBIERA MOROLLÓN, Fernando. Labor density and wages in Spain: Evidence from geographically disaggregated data. **Growth and Change**, v. 49, n. 1, p. 55-70, 2018.

DELLER, Steven; WATSON, Philip. Spatial variations in the relationship between economic diversity and stability. **Applied Economics Letters**, v. 23, n. 7, p. 520-525, 2016.

DE VREYER, Philippe; SPIELVOGEL, Gilles. Spatial externalities between Brazilian municipios and their neighbours. **IAI Discussion Papers**, 2005.

DISSART, Jean Christophe. Regional economic diversity and regional economic stability: research results and agenda. **International Regional Science Review**, v. 26, n. 4, p. 423-446, 2003.

ELHORST, J. Paul et al. Spatial econometrics: from cross-sectional data to spatial panels. **Heidelberg: Springer**, 2014.

GLEASER, Edward L. et al. Growth in cities. **Journal of political economy**, v. 100, n. 6, p. 1126-1152, 1992.

GOODCHILD, Michael F. Scale in GIS: An overview. **Geomorphology**, v. 130, n. 1-2, p. 5-9, 2011.
GOODCHILD, Michael F. et al. Toward spatially integrated social science. **International Regional Science Review**, v. 23, n. 2, p. 139-159, 2000.

HONG, Saheum; XIAO, Yu. The influence of multiple specializations on economic performance in US Metropolitan Areas. **Sustainability**, v. 8, n. 9, p. 963, 2016.

KEINATH JR, William F. The spatial component of the post-industrial society. **Economic geography**, v. 61, n. 3, p. 223-240, 1985.

KELEJIAN, Harry H.; PRUCHA, Ingmar R. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. **The journal of real estate finance and economics**, v. 17, p. 99-121, 1998.

KORT, John R. Regional economic instability and industrial diversification in the US. **Land Economics**, v. 57, n. 4, p. 596-608, 1981.

KRUGMAN, Paul. Increasing returns and economic geography. **Journal of political economy**, v. 99, n. 3, p. 483-499, 1991.

KRUGMAN, Paul. Urban concentration: the role of increasing returns and transport costs. **International Regional Science Review**, v. 19, n. 1-2, p. 5-30, 1996.

KRUGMAN, Paul. What's new about the new economic geography?. **Oxford review of economic policy**, v. 14, n. 2, p. 7-17, 1998.

LESAGE, James P. The theory and practice of spatial econometrics. **University of Toledo. Toledo, Ohio**, v. 28, n. 11, p. 1-39, 1999.

LESAGE, J.; PACE, R. Introduction to spatial econometrics. crc press, boca raton. **Introduction to spatial econometrics**, v. 1, p. 01, 2009.

MALIZIA, Emil E.; KE, Shanzi. The influence of economic diversity on unemployment and stability. **Journal of Regional Science**, v. 33, n. 2, p. 221-235, 1993.

MIZUNO, Keizo; MIZUTANI, Fumitoshi; NAKAYAMA, Noriyoshi. Industrial diversity and metropolitan unemployment rate. **The Annals of Regional Science**, v. 40, p. 157-172, 2006.

NEVES, Mateus de Carvalho *REI*s; CASTRO, Lucas Siqueira de; F*REI*TAS, Carlos Otávio de. O impacto das cooperativas na produção agropecuária brasileira: uma análise econométrica espacial. **Revista de Economia e Sociologia Rural**, v. 57, p. 559-576, 2019.

OPENSHAW, Stan. The modifiable areal unit problem. **Concepts and techniques in modern geography**, 1984.

PORTER, Michael E. The competitive advantage of nations The Free Press. **New York**, v. 564, 1990.

PUGA, Diego; VENABLES, Anthony J. The spread of industry: spatial agglomeration in economic development. **Journal of the Japanese and International Economies**, v. 10, n. 4, p. 440-464, 1996.

RESENDE, Guilherme Mendes. Multiple dimensions of regional economic growth: The Brazilian case, 1991− 2000. **Papers in Regional Science**, v. 90, n. 3, p. 629-662, 2011.

SIEGEL, Paul B.; JOHNSON, Thomas G.; ALWANG, Jeffrey. Regional economic diversity and diversification. **Growth and change**, v. 26, n. 2, p. 261-284, 1995.

SOUZA, Helson Gomes de et al. Uma análise espacial de como o crescimento e a desigualdade afetam a pobreza no Brasil. **Revista de Administração Pública**, v. 55, p. 459-482, 2021.

TOBLER, Waldo R. A computer movie simulating urban growth in the Detroit region. **Economic geography**, v. 46, n. sup1, p. 234-240, 1970.

TRENDLE, Bernard; SHORNEY, Gianna. The effect of industrial diversification on regional economic performance. **Australasian Journal of Regional Studies**, v. 9, n. 3, p. 355-369, 2003.

WAGNER, John E. Regional economic diversity: action, concept, or state of confusion. **Journal of Regional analysis and policy**, v. 30, n. 1100-2016-90005, 2000.

WAGNER, John E.; DELLER, Steven C. Measuring the effects of economic diversity on growth and stability. **Land Economics**, p. 541-556, 1998.

WATSON, Philip; DELLER, Steven. Economic diversity, unemployment and the Great Recession. **The Quarterly Review of Economics and Finance**, v. 64, p. 1-11, 2017.