Title: Data concentration analysis

Area: Economia regional

Autores

1. **Wilfredo Sosa (apresentador)**

   Graduação na “Universidad Nacional de Ingenieria"de Lima (Peru, 1980-1985), Mestrado na Universidade Federal do Rio de Janeiro (1994-1996), Doutorado no IMPA (1996-200), PósDoc na “Universitè Blaise Pascal"de Clermont Ferrand da França (2006) e PósDoc na “Universitat Autonoma de Barcelona"de Barcelona na Espanha (2020). Membro Titular da Academia de Ciencias do Peru (desde 2010). Professor da Universidade Catolica de Brasilia (desde 2011).

2. Jose Antonio de França

   Doutor em Contabilidade. Doutor em Economia. Mestre em Administração. Mestre em Contabilidade. Especialista em Auditoria Interna e Externa. Especialista em Análise de Sistemas. Graduação em Contabilidade. Graduação em Economia. Auditor Independente. Perito Contábil. Professor do Departamento de Ciências Contábeis e Atuariais da Universidade de Brasília (UnB). Líder do Laboratório de Pesquisa e Extensão do terceiro Setor (LPETS) da UnB. Professor credenciado do Programa de Pós Graduação em Governança e inovação em Políticas Públicas da UnB.

3. Fernando Silva Vinhado

   Mestre e Doutor em Economia pela Universidade Católica de Brasília (UCB), pós graduado em gestão financeira e controladoria pela Fundação Getúlio Vargas (FGV) e Bacharel em Ciências Contábeis pela Universidade de São Paulo (USP). Funcionário de carreira no Banco do Brasil há mais de 20 anos, com experiência nas áreas de Gestão de Riscos, Controladoria e Crédito, exercendo atualmente a função de Gerente de Soluções na Gerência de Assessoramento Econômico (Gease). Atuou como professor em instituição de ensino superior ministrando disciplinas nas áreas de economia, finanças e contabilidade e possui publicações em periódicos nacionais e internacionais.

Data Concentration Analysis

20 de julho de 2021

Resumo

In this article we introduce a non-parametric model to analyze the concentration of productive sectors of an economy defined by S productive sectors and R regions at a horizon H of times. We extend the notion of location quotient, from two to three dimensions, adding an horizon time which is divided by periods. The data is a tridimensional database and the analysis consists of two parts: the global part, where we focus on the dynamics of the global participation of each productive sector in the economy at each unit of time, for this we introduce the global participation matrix for each productive sector in the economy; and the local part, where we focus on the dynamics of the local concentration of all sectors in each region at each period, in order to analyze it we introduce, for each region and each period of time, the regional matrix of the location quotient and its concentration index which is nothing else that the norm of the vector contain the eigenvalues of its covariance matrix. Both the global and the local parts, are complemented in order to obtain the concentration analysis of productive sectors.

Keywords:Non-parametric model, Location quotient, Concentration index

JEL subject classification: C02, C23, R10, R11, R12

1 Introduction

The concentration of productive sectors in each region of an economy is a very interesting topic in the regional economy. In this context, we introduce a non-parametric model to analyze the concentration of S productive sectors in R regions at a horizon of H units of time of an economy. According to the literature, there is already a concept called location quotient (LQ)

introduced by Robert Murray Haig in 1920 (for more details see [1]), which was considered as a measure of concentration of a sector s in a region r.

The seminal conception of QL is associated with studies of agglomerations (Haig, 1920) and Florence (1929). These studies, intensified since the beginning of the 20th century, have relied on a quantitative measure to assess the growth and development of the economy in the context of regions. Agglomeration, as referred in the literature, is equivalent to grouping and among the quantitative tools used to measure its effects, the QL model structured in two dimensions has traditionally stood out: economics (sectors of the economy) and geography (regions of the economy). The spatial measurement literature, in the context of regional economics, is part of economic theory and signals that the agglomeration in geography implies a grouping of people, wealth and diffusion of skills that can occur in successive regions, as declared by Marshal (1890, IV , X, 2; IV, X, 3). This grouping can be explained by the laws of integration and agglomeration. These laws declare a tendency for people to group together to combine social and economic forms of coexistence that grow and interact with each other, evolving from independent groups to societies with their respective political systems, as noted by Florence (1929).

In the “economy"dimension, the agglomeration of industry, represented by the production sectors, promotes the appearance of other regional agglomerations, promoting an increase in the level of income, and, consequently, an increase in social well-being, as addressed by Gabe and Abel (2010). But it was Marshal (1890; IV, X, 3-p.162) who initially stated that when an industry chooses a place to set up it is likely to stay there for some time because there are advantages for those who follow it because of the neighborhood specialized. To justify his perception, the author declares "The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously."Several other studies on agglomerations present arguments that support a relevant part of the economic theory applied to regional development. These studies categorize economic activity, as in Tiebout (1962), Alexander (1954) and Thulin (2014), which measure the effect of each sector on the spatial activity of the economy through the LQ. In contrast, studies such as those by North (1955), Gilmer and Keil (1989) and Gabe and Abel (2010) disagree the use of the QL analytical model to measure the effects of agglomerations because they understand that the interaction between members of agglomerations, internally, it can capture benefits that depend on the outside world, and the interaction with other individuals outside the agglomeration may be limited by settlement patterns. This criticism seems to be consistent when the model incorporates some economic variables, such as production, employment and income, because important premises, such as the definition of public politic and periods of a time horizon, cannot be considered in the model with only two dimensions. Previous studies of location and agglomeration are focused on the development of regions such as, localities, districts and municipalities, centered on the specialization of economic, intellectual and equivalent activities, as initially declared by Marshal (1890, IV, X, 2; IV, X, 3) who described the foundations of regional economic thinking for discussing the advantages and benefits of the agglomeration of industry, in a broad way. The advantages are attributed to the proximity among manufacturing and primary exploration of the raw material, production and transportation infrastructure, technology and communication infrastructure. The benefits, as corroborated by Gabe and Abel (2010), are inherent to the overflow of knowledge that inhibits the reduction of production considering the competition between groups of specialized knowledge.

Other scholars in this segment of economic theory have also made efforts to validate some indicator capable of measuring the effects of the agglomeration of an economic activity, in a region, with the purpose of identifying the contribution of the sectors called basic and non-basic for the whole industry. The literature refers to the basic sector as the sector that comprises the specialty of the industry that exports commodities and services; non-basic sector is one that supports the basic industry and serves the region itself. These concepts, as stated in the report of the Division of Research and statistics of New York State of Opportunity (Jun, 2017), are attributed to Haig (1928), and are part of the New York City Regional Plan (1928). In addition to these two concepts, the report also credits Haig (1920) the location quotient model, declared below, as an indicator capable of measuring the contribution of the basic and non-basic sectors of employment in the United States.

$$LQ = \frac{\text{Industry } X's\%\text{Share of jobs in the Regional Economy}}{\text{Industry } X's\%\text{Share of jobs in the U.S. Economy}}$$

From this relationship, the following interpretation follows:

$$LQ = \begin{cases} > & 1 \quad \textit{implies} \quad \text{exports the surplus} \\ = & 1 \quad \textit{implies} \quad \text{in balance} \\ < & 1 \quad \textit{implies} \quad \text{insufficient production} \end{cases}$$

But the literature is not unanimous about the authorship of the LQ and its contribution to explain aspects of the United States' economy, as already mentioned in the previous paragraph. About authorship, in addition to Haig (920), as mentioned in the New York Regional Plan (1928), Thulin (2014) attributes to Florence (1929). Florence (1929-pp. 115-116; 327-328; 435) addresses the issues of location and city and country cycles. She declares that urbanization and the location of a population result in a geographic agglomeration of people and wealth, among others, which may occur in successive regions. For the author, industrial concentration is measured by the economic activity of businesses whose location happens in certain places where there is population density. In this analysis, he understands that it is necessary to know two issues such as: amplitude, extension and form of concentration, as well as the specific moment or place in which the concentration occurs. To this end, it relies on the laws of integration and agglomeration (already mentioned) and highlights, among others, the law of size or agglomeration which states that men tend to unite in groups, and these groups combine increasingly larger social and economic forms. In this context, and in addition, the author (p.328) also suggests a model for measuring the degree of location / concentration of the wool industry in Bradford, as follows:

$$\text{Local concentration} = \left(\frac{A}{C}\right)\left(\frac{B}{D}\right)^{-1}$$

The model variables are described as follows: A is a location with a specified industry; C is a location that brings together all the specified industry; B are all locations where there is a specified industry; D are all locations of all specified industries.

Despite to some controversies evidenced by empirical research, other researches corroborate with the concern to propose a model of agglomeration indicator and suggest that the LQ is credible to assess the concentration and diffusion of regional economic activities. In this context Richardson (1985) declares that the LQ is the appropriate indicator because it is based on the following premises: consumption patterns must be uniform across the region; labor productivity must be constant across regions; local demand must be met, when possible, by production; and the country must be understood as economically sustainable. With these assumptions, as in Haig (1920), Stimson, Stough and Roberts (2006), the conclusion is that if $QL > 1$ the region is self-sustaining or exporting and in that condition local employment is leveraged. In addition to the use in employment statistics, Thulin (2014) highlights that the LQ is not limited to this application, because it can measure other economic activities such as income and wages. It is in this sense that this quotient has been spread as an indicator capable of suggesting levels of concentration of industry activities, as shown by Stimson, Stough and Roberts (2006, p. 107). The authors report to the labor market with an emphasis on employment, as shown in the following equation. If LQ is understood as a parametric measure, it signals increasing levels of performance when $QL \geq 1$, otherwise there is no evidence of agglomeration of specialized activity.

$$QL_{ir} = \left(\frac{E_{ir}}{E_r}\right)\left(\frac{E_{iN}}{E_N}\right)^{-1}$$

The model specification states that: E_{ir} is employment in sector i in region r; E_r is the total employment in region r; E_{iN} is employment in sector i in the reference area ($N =$ national reference); E_N is the total employment in the national reference area. Now, to end the discussion on the classic two-dimensional LQ model, consider an economy with R regions and S productive sectors. Since Marshal (1890), literature has used LQ in the regional economy as an tool of positivist research to measure the degree of concentration of a productive sector in a region. To define the location quotient, as this research proposes, corroborating the seminal studies in two dimensions, a new economic variable observed in each productive sector, by region, is required, according to the model introduced by Haig (1920) and Florence (1929) , as follows:

$$LQ(i,k) = \frac{\left[\frac{V(i,k)}{\sum_{j=1}^{S} V(i,j)}\right]}{\left[\frac{\sum_{l=1}^{R} V(l,k)}{\sum_{l=1}^{R}\left(\sum_{j=1}^{S} V(l,j)\right)}\right]}$$

The model specification suggests that: $V(i,k)$ is the measure of the economic variable observed in region i and in sector k. However, the results presented by this model specification, recurrent in the literature, can lead to misleading inferences when they suggest that the region is an importer or exporter because it has a LQ less than or greater than 1. In this case, the unit is a balance parameter. Misleading interpretation can happen when the result presented by the LQ is the result of a process with less or greater specialization due to the internal use of lesser or greater technology. In the first situation, a region that uses less technology than another, in some sector, will exhibit $QL < 1$. Otherwise, in that region, $QL \geq 1$. In both situations, this technology effect may not mean that, in some sector, whether the region is an importer or an exporter, but that there is a better use of installed capacity with the allocation of resources.

Here, the classical matrix LQ is modified, considering an extra dimension, the time. So the LQ, with three dimensions, is a three dimensional database of location quotients, where each component is a location quotient. So, the first thing this model does is to transform the three dimensional database V into a three dimensional database of location quotients LQ. After it, in the horizon of time H, the model build periods, where each period has P units of time, as follows: The first period $p = 1$ consists of the first P unit of time, and the next period is building eliminating the time p and adding the time $p + P$. So, we have $H - P + 1$ periods. Notice that a period can be seen as a moving window of size P spanning the complete horizon H. So, fixing a region r and the period p, we introduce the regional matrix of location quotients denoted by $RMLQ_r^p$ such that $RMLQ_r^p(s, r, t) = LQ(s, r, t)$ with $s \in \{1, 2, \cdots, S\}$ and $t \in \{p, p + 1, \cdots, p + P - 1\}$. From $RMLQ_r^p$, we introduce the concentration index $CI(r, p)$ of the region r at period p.

The analysis of the concentration of the productive sector is building in two part:

1. The global analysis, where the dynamic of the participation of each sector in the economy is is modulated according to some regulator's strategy along the periods in the horizon. For it, we introduce the global participation matrix GP, which describe the dynamic of the participation of each sector in the economy.

2. The local analysis, where we introduce the concentration index matrix CI,such that $CI(r, p)$ measure the concentration of all productive sectors in the region r at the period p and is obtained from the matrix $RMLQ_r^p$, in order to define the modulation of the global percentage matrix at the final of each period p.

In the section 2, we introduce the three dimensional data of location quotients. In the section 3, we introduce the non parametric methodology. In the section 4 we consider an academical example and in the last section we make a conclusions.

2 The three dimensional database of location quotients

In regional economy, given an economy with S productive sectors and R regions where an economical variable V is observed in a horizon H (for example, the variable $V(s, r, t)$ would be the investment of the sector $s \in \{1, 2, \cdots, S\}$ in the region $r \in \{1, 2, \cdots, R\}$ at time $t \in \{1, 2, \cdots, H\}$).

Before to define the three dimensional data of location quotients, we consider the following notations:

1. $TR(r,t) = \sum_{k=1}^{S} V(k,r,t)$, the aggregation of all sector in the region r at time t. TR is a matrix with size $R \times H$.

2. $TS(s,t) = \sum_{i=1}^{R} V(s,i,t)$ the aggregation of all regions in the sector s at time t. TS is a matrix with size $S \times H$

3. $T(t) = \sum_{r=1}^{R} TR(r,t) = \sum_{s=1}^{S} TS(s,t)$, the total aggregation of the economy at time t. T is a vector in $\mathbb{R}^H$.

4. $LP(s,r,t) = [V(s,r,t)/TR(r,t)]$ is called the local participation of the sector s in the region r at time t. So LP is a three dimensional database with size $S \times R \times H$.

5. $GP(s,t) = [TS(s,t)/T(t)]$ is called the global participation of the sector s at time t. So GP is a matrix with size $S \times H$

The location quotient is defined by

$$LQ(s,r,t) := \left[\frac{\left[\frac{V(s,r,t)}{TR(r,t)}\right]}{\left[\frac{TS(s,t)}{T(t)}\right]} \right] = \left[\frac{LP(s,r,t)}{GP(s,t)} \right]$$

Thus, the location quotient is an operator that transforms the three dimensional database V that contains the data of the observed variable of the economy, into a three dimensional database LQ which each component is a location quotient. On the other hand, each component of the LQ, is nothing else than the quotient between the local participation of a productive sector in the a region at a time over the global participation of the same productive sector in the economy at the same time.

3 The non parametric methodology

Our methodology is non-parametric because it only uses the three dimensional database V to create, tools for our analysis, for instance:

1. The matrix TS.

2. The matrix GP.

3. The three dimensional database LP.

4. The three dimensional database LQ.

For the analysis of the concentration of the productive sectors, we consider three assumptions:

1. The regulator of the economy, define the participation of each productive sector in the economy. We denote, this participation of each productive sector in the economy, by $\bar{s}$. It will called the strategy objective of the economy and it is defined a priori exogenously.

2. The regulator has packages of public politics, either for increase or decrease the participation of each productive sector in the economy and for define the size of the economy, at the end of each period, in order to modulate the participation of each productive sector in the market. When the regulator executes a package of public politics geared towards meeting its strategy objective of the economy, the regulator observes for a period whether the public politic was successful or not. These public politics are part of the regulation of the market.

3. When $|GP(s,t) - \bar{s}(s)| > \epsilon$ for $t \geq P$, then the sector s need to redefine the public politic for the period $p = t - P + 2$. The parameter $\epsilon \in (0,1)$ is an acceptation error to measure the package of public politics for the sector s.

3.1 The global analysis

By definition $GP(s,t) = \left[\frac{TS(s,t)}{T(t)}\right]$. So,

$$\sum_{s=1}^{S} GP(s,t) = \sum_{s=1}^{S} \left[\frac{TS(s,t)}{T(t)}\right] = \left[\frac{\sum_{s=1}^{S} TS(s,t)}{T(t)}\right] = 1$$

It is, $GP(s,t)$ represent the global participation of the sector s at time t in the economy. So, the matrix GP contains the dynamic of the global participation of each sector s along the horizon time H.

The main tool for this global analysis is the matrix GP, because for each time t in the horizon H, by definition, the column t of the matrix GP contain the global participation of each sector s in the economy. It is, the regulator has the matrix GP and its packages of public politics in order to modulate the global participation of each sector in the market.

We have the following behaviors for each curve of sector $s \in \{1, 2, \cdots, S\}$, $\{GP(s,t)\}$ at period a $p \in \{1, 2, \cdots, H-P+1\}$ ($t \in \{p, p+1, \cdots, p+P-1\}$):

1. The behavior, is non decrease.

2. The behavior, is decrease.

3. The behavior, is non increase.

4. The behavior, is increase.

5. The behavior, is stable.

Each behavior, has the following interpretation:

1. When it is non decrease and $GP(s,p+P-1)-\bar{s}(s) > \epsilon$ at the period p, the current package of the public politics is no working. New package is required for the period $p+1$.

2. When it is decrease and $GP(s,p+P-1)-\bar{s}(s) > \epsilon$ at the period p, the current package of public politics is working.

3. When it is non increase and $\bar{s}(s)-GP(s,p+P-1) > \epsilon$ at the period p, the current package of public politics is no working. New package is required for the period $p+1$.

4. When it is increase and $\bar{s}(s)-GP(s,p+P-1) < \epsilon$ at the period p, the current package of public politics is working.

5. When it is stable and $|\bar{s}(s)-GP(s,p+P-1)| \leq \epsilon$ at the period p, the current package of public politics is working.

The global analysis is not only for regulate respect to the strategy objective of the market. The globally analysis is also to regulate the size of the market, keeping the strategy objective of the market.

So, when the regulator establish increase or decrease or maintain the size of the market, keeping the strategy objective of the market, he has its packages of public politics and the matrix $TS(s,t)$ which gives the dynamic of the aggregate production of sector s in the economy, in order to regulate, globally, the market.

Matrices GP and TS give us, only, global behaviors of productive sectors in the economy. Both matrices are building using the three dimensional data base V.

3.2 The Local analysis

In addition to matrices GP and TS, the three dimensional database V, also it is used to builds the three dimensional database LP and the three dimensional database LQ. From the three dimensional database LQ, we define the regional matrices of of location quotients $RMLQ_r^p$, for each region $r \in \{1, 2, \cdots, R\}$ and each period $p \in \{1, 2, \cdots, H - P + 1\}$ as follows

Definition 1. *For each $r \in \{1, 2, \cdots, R\}$ and each $p \in \{1, 2, \cdots, H-P+1\}$, we define and denote the regional matrix of location quotients by*

$$RMLQ_r^p(s, t) = LQ(s, r, t) \ \forall s \in \{1, 2, \cdots, S\} \ \forall t \in \{p, p+1, \cdots, p+P-1\}$$

So, the matrix $RMLQ_r^p$ has S rows and P columns and its components are location quotients.

Now, we introduce some properties of regional matrices of location quotients.

Proposition 1. *For each matrix $RMLQ_r^p$ with $r \in \{1, 2, \cdots, R\}$ and $p \in \{1, 2, \cdots, H - P + 1\}$ and each t in the period p, the following statements are equivalent:*

1. *There exists $s_1 \in \{1, 2, \cdots S\}$ such that $LQ(s_1, r, t) > 1$.*
2. *There exists $s_2 \in \{1, 2, \cdots S\}$ such that $LQ(s_1, r, t) < 1$.*
3. *The matrix $RMLQ_r^p$ is no constant.*
4. *The covariance matrix of $RMLQ_r^p$ has at least one eigenvalue strictly positive.*
5. *The covariance matrix of $RMLQ_r^p$ has at least one component no null.*

As a consequence we have the following corollary.

Corollary 1. *The following statements are equivalent.*

1. *$RMQLR_r^p$ is a constant matrix (each component is one).*
2. *The covariance matrix of $RMQL_r^p$ is null (each component is zero).*
3. *All the eigenvalues of the covariance matrix of $RMQL_r^p$ are zeros.*
4. *The norm of the vector contains all the eigenvalue of the covariance matrix of $RMQL_r^p$ is zero.*

Let CV_r^p be the covariance matrix of $RMQL_r^p$. Since CV_r^p is symmetric and semi definite positive, it has real eigenvalues and eigenvectors. The interpretation of each eigenvalue λ is as follows: if m is the mean of all rows of the $RMLQ_r^p$ matrix and $\pounds$ is the line that passes through m with direction of the eigenvector v associated with λ, then λ represent the dispersion of the projections of the rows of $RMLQ_r^p$ over $\pounds$. So, all the eigenvalues represent the dispersion/concentration of the rows with respect to m. Then the norm of the vector, that contains all the eigenvalues of the matrix CV_r^p, denoted by $CI(r,p)$ is a measure of the concentration of all productive sectors in the region r at period p. So, we define the matrix of concentration index as follow:

Definition 2. *Given the three dimensional database V, we define its concentration indexes matrix as CI such that each component (r,p) is the norm of the vector contain all the eigenvalues of the covariance matrix of the matrix $RMLQ_r^p$.*

This result establishes an interesting behavior of the rows of regional matrix of location quotients. Here, we supposes that $H = +\infty$ (here $t, p \in \{1, 2, \cdots, +\infty\}$).

Theorem 1. *Let $\{CI(r,p)\}$ be a sequence with $r \in \{1, 2, \cdots, R\}$ fixed. The following statements are equivalent:*

1. *$\{CI(r,p)\}$ converges to zero.*
2. *$\{LQ(s,r,t)\}$ converges to one, for each $s \in \{1, 2, \cdots, S\}$.*

The next result is a direct consequence of previous results.

Corollary 2. *For each matrix $RMLQ_r^p$ with $r \in \{1, 2, \cdots, R\}$ and $p \in \{1, 2, \cdots, H-P+1\}$. $CI(r,p)$ is sufficiently close to zero if and only if each component of $RMQL_r^p$ is sufficiently close to one.*

According to the global analysis, for each period p at time $t = p+P-1$, for each productive sector s we have 5 possibilities. So, we introduce some rules in order to regulate the market according to the structure of the market.

1. Let s be a productive sector. If $|GP(s,t) - \bar{s}(s)| > \epsilon$ and the behavior of the sequence $\{GP(s,t)\}$ in the period p is one of the first three possibilities, then those regions have the highest $CI(r,p)$ and the difference $GP(s,t) - LP(s,r,t) < 0$, then the sector s in those regions r need to decrease its participation, for regulate the market in order to guarantee the structure of the market.

2. Let s be a productive sector. If $|GP(s,t) - \bar{s}(s)| > \epsilon$ and the behavior of the sequence $\{GP(s,t)\}$ in the period p is one of the last three possibilities, then those regions have the highest $CI(r,p)$ and the difference

$GP(s,t) - LP(s,r,t) > 0$, then the sector s in those regions r need to increase its participation, for regulate the market in order to guarantee the structure of the market.

4 Academical exercise

This section consists of an academic example for an economy with three productive sectors, three regions and two time periods ($P = 2$), defined as ($T-P+1 = 2$). In this context, the regulator of the market define the politics he wish to conduct as well as the structure of the economy. But the global percentages of productive sectors depend on regional performance, which is signaled by the local percentages that define how the economy produces in each region. Thus, once the global percentages are defined, the regulator implement them, follow them and monitor them to ensure that the objectives of each politic planned by the regulator are met.

The economy in this academic example is as follows.

(a) Geography dimension: three regions (r_1, r_2, r_3).

(b) Economy dimension: three productive sectors (s_1, s_2, s_3).

(c) Time dimension: Horizon with three years (t_1, t_2, t_3).

The horizon is divided into periods, each period consists of two years, so we have two periods (p_1, p_2).

The structure of the market, defined by the regulator is as follows:

(a) Sector s_1: 35% of participation.

(b) Sector s_2: 20% of participation.

(c) Sector s_3: 45% of participation.

4.1 Concentration analysis

The analysis is make by period. It is, the regulator establish a package of public politics for regulate the three productive sectors in the first period, in order to develop each region. The answer of its package of public politics is as follows:

Region	Sector	t1	t2	Region	Sector	t1	t2	Region	Sector	t1	t2
r1	s1	700	1450	r2	s1	300	300	r3	s1	400	410
r1	s2	1000	1600	r2	s2	450	450	r3	s2	600	550
r1	s3	1735	2200	r2	s3	750	750	r3	s3	1000	930

Tabela 1: First period

4.2 Global analysis in the first period

The global analysis is make using the global percentage matrix GP. SM denote the structure of the market and $\widehat{GP}$ is a matrix such that each column is the difference between the respective column of GP and SM.

$$GP = \begin{bmatrix} 0.20 & 0.25 \\ 0.30 & 0.30 \\ 0.50 & 0.45 \end{bmatrix}, SM = \begin{bmatrix} 35 \\ 20 \\ 45 \end{bmatrix}, \widehat{GP} = \begin{bmatrix} -0.15 & -0.1 \\ 0.1 & 0.1 \\ 0.5 & 0 \end{bmatrix}$$

In the first year, sectors 2 and 3 benefited from the public policy package, while the first sector did not respond to that package. In the second year, the first sector grows, the second maintains the percentage of the first year and the third sector decreases. The sector with problems with the packages of public politics are $s = 1$ and $s = 2$. The regulator should continue to encourage the first sector, reducing the percentage of participation of the second and third sectors, to achieve the desired market structure. In other words, it must make an adjustment to the package of public policies for the second period.

4.3 Local analysis in the first period

To know in which region the public policy package was not efficient, we first need to calculate the location quotient tensor, then we need the regional location quotient matrices, which are in the next table.

Sector	t1	t2	Sector	t1	t2	Sector	t1	t2
s1	1.0095	1.1048	s1	0.9907	0.800	s1	0.9907	0.8677
s2	0.9848	1.0127	s2	1.0149	0.9969	s2	1.0149	0.9670
s3	1.0051	0.9331	s3	0.9950	1.1134	s3	0.9950	1.0957

Tabela 2: Regional matrix of location quotient

The CI matrix for the first period is $CI = \begin{bmatrix} 0.00176 \\ 0.05785 \\ 0.00168 \end{bmatrix}$. Thus, the highest concentration index is that of the second region. It is, the adjustment need to be in the region $r2$. Now, observing the matrix $RMLQ_2^1$, we have that the adjustment is to increase the sector $s1$. The adjustment packages of public politics for the second period is to incentives the sector $s1$ in the region $r2$.

4.4 Global analysis for the second period

The global analysis is make using the global percentage matrix

$$GP = \begin{bmatrix} 0.20 & 0.25 & 0.35 \\ 0.30 & 0.30 & 0.20 \\ 0.50 & 0.45 & 0.45 \end{bmatrix}, \widehat{GP} = \begin{bmatrix} -0.15 & -0.1 & 0 \\ 0.1 & 0.1 & 0 \\ 0.5 & 0 & 0 \end{bmatrix}$$

In the third year the regulator achieves the desired market structure

4.5 Local analysis in the second period

To understand the reaction of the market to the public policy packages, we first need to calculate the location quotient tensor, which which is the next table.

t1	t2	t3	t1	t2	t3	t1	t2	t3
1.0095	1.1048	1.2733	0.9907	0.800	0.5709	0.9907	0.8677	0.5014
0.9848	1.0127	0.7948	1.0149	0.9969	1.5066	1.0149	0.9670	1.1876
1.0051	0.9331	0.8782	0.9950	1.1134	1.1098	0.9950	1.0957	1.3048

Tabela 3: Regional matrices of location quotients

note that the values in the time $t3$ no change the behavior of the location quotients in $t1$ and $t2$.

Respect to the CI matrix, which is: $CI = \begin{bmatrix} 0.00176 & 0.00184 \\ 0.05785 & 0.02577 \\ 0.00168 & 0.21461 \end{bmatrix}$, holds the same. In other words, the future no changes the past.

Conclusion

This methodology is non parametric, because use only the data tensor and transform it in a location quotients tensor. The global analysis is building with the Global percentage matrix, which describe the dynamic of the composition of the market for each productive sector along the horizon H. In order to see the size of the market, the regulator has dynamic of the aggregate by sector along the horizon. The local analysis is used for define at the end of each period, the adjustment of the package of public politics in a region, where the current package has difficulties. For it the regulator use the concentration index matrix CI and the respective regional matrix of location quotients.

Acknowledgments

The second author was supported in part by Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF) by the grant 0193.001695/2017 and PDE 05/2018. This research was carried out, during (the state of alert in Catalonia) a visit, of the second author to the Centre de Recerca Matemática (CRM), in the framework of the Research in pairs call in 2020. The CRM is a paradise for research, the author appreciates the hospitality and all the support received from CRM.

Referências

[1] Robert Murray Haig. Location Quotients: A statewid and Regional Analysis Division of Resarch and Statistics of New York State. (2017), Jun./2017. *Disponível em: https://labor.ny.gov/stats/PDFs/Location-Quotients-A-Statewide-and-Regional-Analysis.pdf.*.

[2] Alexander, John W. The Basic-Nonbasic Concept of Urban Economic Functions. *Economic Geography* V.30, N. 3, pp. 246-261 (Jul., 1954).

[3] Anselin, Luc; Varga Attila; Acs, Zoltan. Local Geographic Spillovers between University Research and High Technology Innovations. *Journal of Urban Economics* v.42, pp. 442-448 (1997).

[4] Antonelli, C. Localised Technological Change. Towards the Economics of Complexity. *Routledge, New York* (2008).

[5] W. Mark Brown, Darren M. Scott. Human capital location choice:Accounting for amenities and thick labor markets. *Journal of Regional Science* v. 52. Nr. 5, pp. 787-808, (2012).

[6] Carleial, Liana. Economia Regional e Urbana. *Ipea Publication* (pp. 113-139), (2011).

[7] Todd Gabe and R. Jaison Abel. Agglomeration of Knowledge. *Urban Studies Journal* DOI: 10.1177/0042098010371988 (2010).

[8] P. Cooke, M. Uranga and G. Etxebarria Regional innovation systems: Institutional and organizational dimensions. *Research Policy* v. 26, 475-491, (1997).

[9] P. S. Florence. Economic research and industrial policy. *The Economic Journal* V. 47, n. 188, pp. 621-641. (1937).

[10] P. S. Florence. The Statistical Method in Economics and Political Science: A treatise on the quantitative and institutional approach to social and industrial problems. *Harcourt, New York.* (1929).

[11] Faggian Alessandra and McCann Philip Human capital, graduate migration and innovation in British regions. *Cambridge Journal of Economics* v. 33, nr. 2, pp.317-333. https://doi.org/10.1093/cje/ben042. (2009).

5 Appendix: Elements of Linear Algebra

Let $\mathbb{R}$ be the set of real numbers, $\mathbb{R}^n$ denote the vectorial euclidian space of dimension $n \in \mathbb{N}$, $\mathbb{R}^{n\times m}$ denote the space of matrices with n rows and m columns (each element of $\mathbb{R}^n$ is a matrix, with n rows and one column). For each $A \in \mathbb{R}^{n\times m}$, A' denote the transpose of A (the transpose, transform every rows of matrix A in a column of matrix A', and conversely). The product of two matrices $A \in \mathbb{R}^{n\times p}$ and $B \in \mathbb{R}^{q\times m}$ is well defined if and only if $p = q$. Moreover, the product AB has size $n \times m$. The norm of a vector $x \in \mathbb{R}^n$ is defined and denoted as $\|x\| = (x'x)^{1/2}$. A matrix $A \in \mathbb{R}^{n\times n}$ is called symmetric if $A = A'$. A matrix $A \in \mathbb{R}^{n\times n}$ is called semi definite if $x'Ax \geq 0\ \forall x \in \mathbb{R}^n$. $\lambda \in \mathbb{R}$ is called an eigenvalue of a matrix $A \in \mathbb{R}^{n\times n}$ if there exists a non null vector x ($\|x\| \neq 0$) such that $Ax = \lambda x$ (x is called the eigenvector associated to the eigenvalue λ). The norm of a matrix $A \in \mathbb{R}^{n\times n}$ is defined by $\|A\| = \max_{\|x\|=1} x'Ax$. It is well known that $\|A\| = 0$ if and only if A is the null matrix (all its components are zeroes). In the concentration analysis, some ideas coming from the linear algebra, which can find in any classical book, will be used, for instance:

1. Every matrix $A \in \mathbb{R}^{m\times n}$ has an associated positive semi defined symmetric matrix (covariance matrix). We denote by A_i its i-th row and a is the mean of all rows, then $(1/m)B^TB$ is the covariance matrix of A, when the i-th rows of $B \in \mathbb{R}^{n\times n}$ is $(A_i - a)$.

2. Every positive semi defined symmetric matrix only has nonnegative eigenvalues.

3. If each one of the eigenvalues of symmetric matrix is zero, then the matrix is null (its components are zeroes).

4. If the covariance matrix of a matrix A is null, then the matrix A is constant (all components are equals).

5. Every sample of m data in R^P, if each data is the row of a matrix A and B is its covariance matrix, then the norm of the vector containing all the eigenvalues of matrix B represents a measure of the concentration/dispersion of the data sample with respect to its median.

Demonstração. of Proposition 1 We only proof 1) implies 2), because the others ones are exercises of Linear Algebra. If there is s_1 such that $LQ(s_1, r, t) > 1$. By contradiction, suppose that $LQ(s, r, t) \geq 1\ \forall s \in \{1, 2, \cdot, S\}$. Then

$$\begin{array}{rcl}
\left[\frac{V(s,r,t)}{TR(r,t)}\right] & \geq & \left[\frac{TS(s,t)}{T(t)}\right]\ \forall s \in \{1, 2, \cdot, S\} \\
\sum_{s=1}^{S}\left[\frac{V(s,r,t)}{TR(r,t)}\right] & > & \sum_{s=1}^{n}\left[\frac{TS(s,t)}{T(t)}\right]\ \textit{because}\ LQ(s_1, r, t) > 1 \\
\left[\frac{\sum_{s=1}^{S} V(s,r,t)}{TR(r,t)}\right] & > & \left[\frac{\sum_{s=1}^{n} TS(s,t)}{T(t)}\right] \\
1 = \left[\frac{TR(r,t)}{TR(r,t)}\right] & > & \left[\frac{T(t)}{T(t)}\right] = 1
\end{array}$$

The last inequality is a contradiction. It implies that 1) implies 2). □

Now, fixing a region $r \in \{1, 2, \cdots, R\}$ and a period $p \in \{1, 2, \cdots, H - P + 1\}$, according to the Linear Algebra, we have a matrix with S rows and P columns, called the regional matrix of location quotients, denoted by $RMLQ_r^p$. This matrix generate the concentration index $CI(r, p)$ of the region r at period p, where $CI(r, p)$ is the norm of the vector contains all the eigenvalues of the covariance matrix of the matrix $RMLQ_r^p$.

Demonstração. of Theorem 1 From the Corollary 1, when a $RMLQ$ is a constant matrix (each component is one) the respective concentration index is zero. On the other hand, the convergence of $\{CI(r, p)\}$ to zero tells us that the characteristic polynomials of order P are converging to the characteristic polynomial of a constant matrix of components equal to one. Furthermore, the convergence of the polynomials is uniform. Thus, we have the convergence of each component of the $RMLQ_r^p$ matrix to one. Reciprocally, the convergence of each $RMLQ_r^p$ component to one implies $CI(r, p)$ converges to zero. □