RESEARCH ARTICLE

# Dynamics in the Fitness-Income plane: Brazilian states vs World countries

Felipe G. Operti[1,2], Emanuele Pugliese[3], José S. Andrade Jr.[1], Luciano Pietronero[2,3,4], Andrea Gabrielli[3]*

**1** Departamento de Física, Universidade Federal do Ceará, 60451-970, Fortaleza, Ceará, Brazil,
**2** Department of Physics, Sapienza University of Rome, 00185, Rome, Italy, **3** Istituto dei Sistemi Complessi (ISC) - CNR, UoS Sapienza, Dipartimento di Fisica, Università Sapienza, P.le Aldo Moro 5, 00185 - Rome, Italy, **4** International Finance Corporation, World Bank Group, 1818 H Street, 20433 - NW Washington DC - United States of America

* andrea.gabrielli@roma1.infn.it

## Abstract

In this paper we introduce a novel algorithm, called *Exogenous Fitness*, to calculate the Fitness of subnational entities and we apply it to the states of Brazil. In the last decade, several indices were introduced to measure the competitiveness of countries by looking at the complexity of their export basket. Tacchella *et al* (2012) developed a non-monetary metric called Fitness. In this paper, after an overview about Brazil as a whole and the comparison with the other BRIC countries, we introduce a new methodology based on the Fitness algorithm, called Exogenous Fitness. Combining the results with the Gross Domestic Product *per capita* ($GDP_p$), we look at the dynamics of the Brazilian states in the Fitness-Income plane. Two regimes are distinguishable: one with high predictability and the other with low predictability, showing a deep analogy with the heterogeneous dynamics of the World countries. Furthermore, we compare the ranking of the Brazilian states according to the Exogenous Fitness with the ranking obtained through two other techniques, namely *Endogenous Fitness* and *Economic Complexity Index*.

## Introduction

Large countries are often characterized by a strong internal heterogeneity between richer regions and poorer hierarchical regions. Just think to the difference between the GDP *per capita* ($GDP_p$) of the states of New York and Mississippi in the US [1], or the difference between the states of Kerala and Bihar in India [2], or between the unexplored forest of Amazon and the modern state of São Paulo in Brazil [3]. While the recent literature on Economic Complexity focused on countries [4–7], we believe that there are two very strong reasons to extend the scope of the analysis to the subnational level.

The first reason is purely academic. Indeed, sharp differences in economic outcomes in a uniform institutional area—with common cultural background and free movement of workers—are both a theoretical puzzle for traditional economics and an empirical opportunity for the Economic Complexity field. Indeed, the analysis of subnational entities competing on an even

playing field is the perfect experimental setup to identify the role of organizational and technical capabilities with respect to more traditional economic factors of analysis. In this paper we will analyze the case of Brazil, to see if the capabilities driven dynamics of a country is replicated at a smaller scale.

The second reason is to improve economic forecasting. Indeed Economic Complexity has been proved to be very effective in forecasting the economic performances of countries [8]. An understanding of subnational entities could give however more accuracy and more detail. It is clear for example that future $GDP_p$ growth of Brazil will depend not only on further growth of the Southern industrial core, but on the convergence of the other regions. This is crucial both to correctly forecast aggregate Brazil $GDP_p$ growth and to address the vast internal inequality of the country.

Brazil or, officially, the *Federative Republic of Brazil* is the ninth World economy in the GDP ranking of the year 2015 [9]. Its population is equivalent to 2.81% of the total World population [10] and its large area (8.515.767,049 $km^2$), divided by its twenty-seven Federative Units [11], make it the fifth largest country of the World [12]. The political and administrative organization of Brazil is hierarchically organized in a sequence of geopolitical structures: Union, states, Federal District and Counties. Each one is autonomous and organized according with the division of powers: legislative, executive, and judiciary. Due to the deep inequalities, but also for the good perspectives of growth, Brazil and the others Latin American countries were often a focus of economic development analysis during the last century [13–17].

Economists usually focus on monetary based indices to analyze economies such as the GDP. However, GDP alone, as shown by different studies [4–7, 18], does not provide deep information about the perspective of growth and development of World countries. Several studies tried to gain information on the unobservable characteristics of countries by looking at stock indices to exploit the "wisdom of the crowd" [19, 20]. In order to gain a direct measure of the country capabilities, the last decade has been marked by a line of research of new indices inspired by the science of complex systems, able to better describe and explain the large scale World economy [4–7, 21–23] and to estimate global and regional inequalities [24, 25].

In this respect, different authors recently introduced two indices: Economy Complexity Index (ECI) [4] and Fitness [5]. Furthermore, Cristelli *et al* [7], through a novel method called *Selective Predictability Scheme* (SPS), showed that the comparison between $GDP_p$ and Fitness provides a highly performing forecasting tool for several countries.

In this paper, we first present an overview of Brazil as a whole from the point of view of the Economic Complexity approach. In this context we compare its export basket and its Fitness with the ones of the BRIC group of countries (Brazil, Russia, India, and China) [26].

Then, we focus on the comparative study of the economies of the single Brazilian states. Based on the "classical" Fitness algorithm, we introduce a new methodology, called Exogenous Fitness, able to measure the Fitness of subnational entities, and we apply it to the states of Brazil. In analogy with what was proposed in [7], we analyze the coevolution of $GDP_p$ and Fitness studying in this way the predictability of the economic growth of the Brazilian states.

Furthermore, we compare the Exogenous Fitness with: (i) the (Endogenous) Fitness -i.e, the natural application of the "classical" Fitness algorithm to the subnational entities of a country-; (ii) the results published by the *Dataviva* platform (an application of the ECI algorithm) [27].

The paper is structured as follows: first we introduce the methods and we provide an overview about Brazil. Then, we show the results of the Exogenous Fitness applied to the Brazilian states and the comparisons with the other techniques. Finally, we conclude with a general discussion about the implications of the results with respect to both points of view of scientific community and policy makers. In the Appendix A, we describe in detail the used database.

## Methods

In this section we describe the algorithms and methods involved in the calculation of the states and countries Fitness coupled to the Complexity of exported products.

### Revealed Comparative Advantage (RCA)

The Revealed Comparative Advantage (RCA) [28] is a quantitative criterion to assess the relative advantages of a country, or, in this case, of a Brazilian state, in the export of certain products compared to the average export of those products. Defining $q_{sp}$ as the flow of the export (in US dollars) of the product $p$ by the state $s$ (see the section *Database* for the data origin), the RCA is defined as:

$$RCA_{sp} = \frac{\dfrac{q_{sp}}{\sum_{p'} q_{sp'}}}{\dfrac{\sum_{s'} q_{s'p}}{\sum_{s'p'} q_{s'p'}}}.$$

(1)

Therefore, it is the ratio between the share of the export of product $p$ with respect to the total export of State $s$ divided by the share of the export of product $p$ with respect to the total Brazilian export.

From the calculation of the RCA for each state-product pair, we build the binary state-product matrix $M_{sp}$. We consider the state $s$ an exporter of a product $p$, if $RCA_{sp} \geq 1$ and, consequently, we set $M_{sp} = 1$. On the contrary, if $RCA_{sp} < 1$, we set $M_{sp} = 0$.

An analogous criterion is used to define the World countries-products matrix $M_{cp}$ (see the section *Database*). This binary matrix shows which country has a comparative advantage in a certain product with respect to the World average [29, 30].

### (Endogenous) Fitness

Recently, different studies have shown the economic relevance of the diversification of the export basket for the competitiveness of a country [4, 6]. The matrix $M$ shows a substantial nested structure highlighted by a strong triangularity, which can be interpreted in the following way: each country approximately exports all the possible products it has the capabilities to produce [5].

Here, considering the geographic size of Brazil and its federal structure, we assume that the same concept is also valid to understand the development and growth of its states. In this framework, we apply the Fitness algorithm to the states-products matrix of elements $M_{sp}$ above defined [5], a statistical approach based on non linear maps coupling Fitness of states and Complexity of Products, to compare Brazilian states. The (Endogenous) Fitness algorithm is defined by the following iterative equations [6]:

$$\begin{cases} \tilde{F}_s^{(n)} = \Sigma_p M_{sp} Q_p^{B\ (n-1)} \\ \\ \tilde{Q}_p^{B\ (n)} = \dfrac{1}{\sum_s M_{sp} \dfrac{1}{F_s^{(n-1)}}} \end{cases} \rightarrow \begin{cases} F_s^{(n)} = \dfrac{\tilde{F}_s^{(n)}}{\langle \tilde{F}_s^{(n)} \rangle_s} \\ \\ Q_p^{B\ (n)} = \dfrac{\tilde{Q}_p^{B\ (n)}}{\langle \tilde{Q}_p^{B\ (n)} \rangle_p}. \end{cases}$$

(2)

The elements $M_{sp}$ are the elements of the previously discussed binary states-products matrix. $\tilde{F}_s^{(n)}$ and $\tilde{Q}_s^{B\ (n)}$ are intermediate variables which are subsequently normalized at each iteration.

The initial conditions satisfy the relations: $\tilde{F}_s^{(0)} = C$ and $\tilde{Q}_p^{B\ (0)} = C$, where we assume $C = 1$ for each state $s$ and for each product $p$ [6].

At each iteration of the algorithm, the Fitness of each state is proportional to the sum of its exported products weighted by their Complexity stressing the importance of having at the same time both a diversified export basket and the most complex possible products in it. The formula for the Complexity of a product is motivated by the following argument: the more the exporters of a product and the smaller their Fitness, the less its expected from the Complexity. In this manner, a state with low Fitness abruptly influences the Complexity of all the products it exports [6]. Therefore, an highly Complex product is made only by few countries/states with high Fitness, while a little Complex product can be made by all the countries/states, both with high and low Fitness. The stability and robustness of the algorithm has been studied in [6, 31] and the Fitness ranking of the states and the Complexity ranking of the products is unambiguously defined after a large enough number of iterations.

Fig 1 shows the matrix $M_{sp}$ of the year 2015, by ordering the states according to the Fitness (the upper the higher complexity), and the products according to the Complexity (the more right the higher the complexity).

In Fig 2 we show the products *spectroscopy* [32] of the years 2005 (dashed lines) and 2015 (filled colors) for few Brazilian states such as: São Paulo, Paraná, Ceará, and Roraima. The spectroscopy is a graphic representation of the export volume (in US Dollars) of a state for each product with $M_{sp} = 1$ ordered at increasing Complexity from left to right [32]. We subsequently group the products (10 for bin) and we summed the export volumes of each product inside each bin. The spectroscopy allows to compare the diversification and the Complexity of the exportation of the states. The figure shows the spectroscopy of high Fitness states such as São Paulo (diversified all along the Complexity spectrum) and Paraná (with a clear peak on medium-high Complexity products), a middle rank state such as Ceará and a low Fitness state such as Roraima (with few low Complexity exports). From the figure, it emerges that a very developed state such as São Paulo has a high flow of exports for a very diversified number of products with a bias towards the high Complexity ones. Paraná has a high peak in several complex products, while Roraima has only one peak in the less complex products. Ceará is a middle ground between the two.
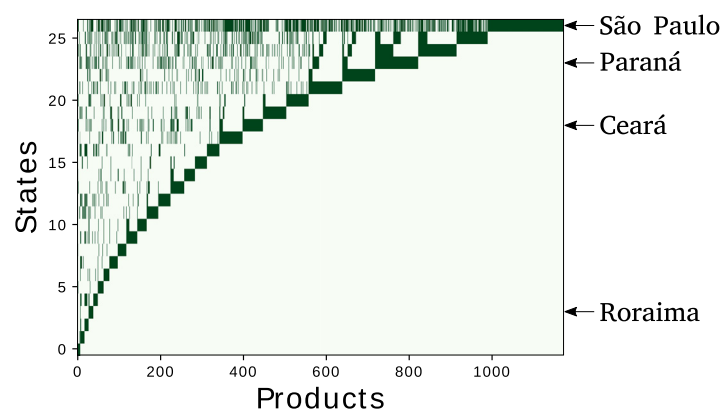


**Fig 1. The binary matrix $M_{sp}$ of the year 2015.** Each row of the matrix represents a Brazilian state. States are ordered in terms of their Fitness from the smallest value (row 0) to the largest one (row 26). Analogously columns represent Products ordered in terms of their Complexity from the smallest value (column 0) to the largest one (column 1172). The matrix elements $M_{sp}$ are drawn in dark green and the others in white. In the figure we highlight high Fitness states such as São Paulo and Paraná, a middle rank State such as Ceará and a low Fitness state such as Roraima.

**Fig 2. Products spectroscopy of the years 2005 (dotted lines) and 2015 (filled colors) of the states: a) São Paulo, b) Para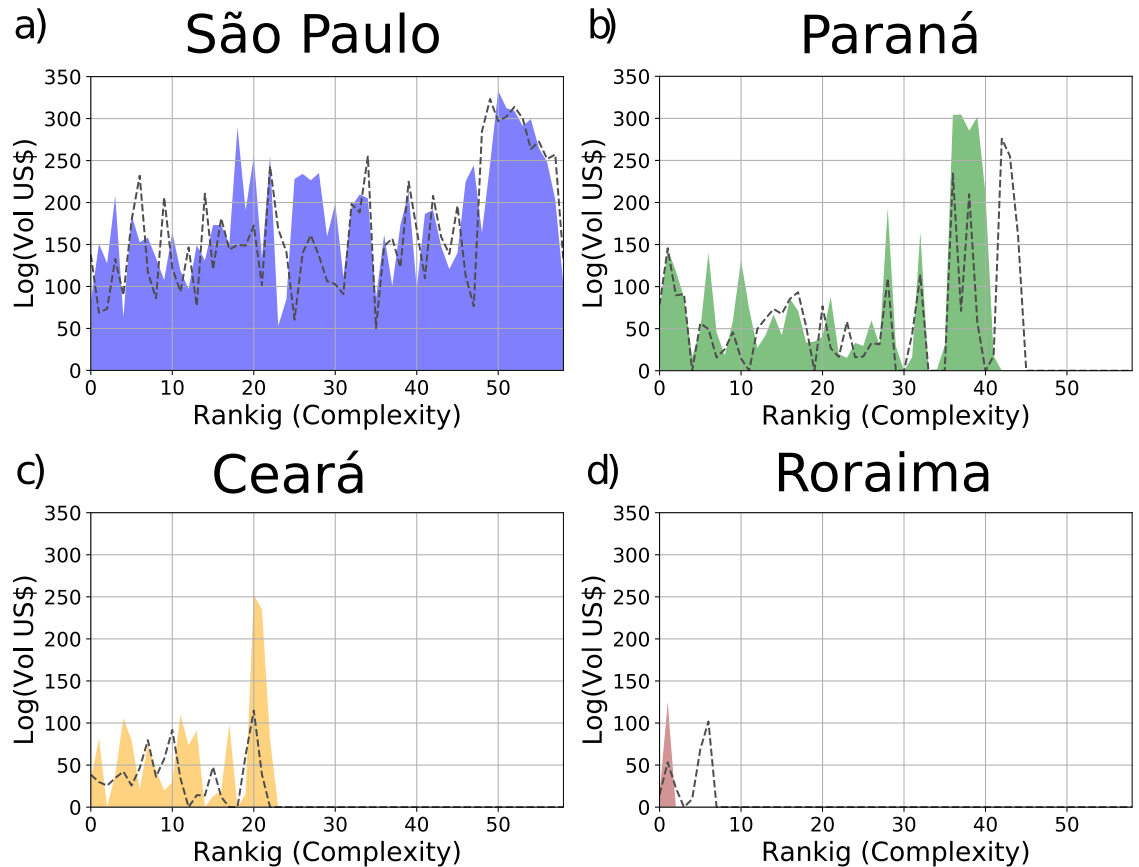ná, c) Ceará, and d) Roraima.** The figures show the export volume (in US Dollars) of those states for each product with $M_{cp} = 1$ ordered according to their Complexity. Products are grouped in bins of 10 and the export volume in each bin are summed up.

### Exogenous Fitness

Here, we define the new Exogenous Fitness algorithm, an innovative method to calculate the Fitness of subnational entities of a country grounded on the measure of the products Complexity from the World-wide trade network. Exogenous Fitness is a coherent extension of the "classical" Fitness algorithm [6], with the assumption of an obvious concept: products have an intrinsic Complexity, reflected by the trade on the global World scale by all countries, while the trade from the regions of a single country may not represent well such intrinsic Complexity as it can be affected by local biases. In particular if we consider only Brazil to define the Complexity of the exported products, we can introduce local economic biases in its measure related to the peculiar features of Brazil economy. Indeed, as shown in Fig 1, there is a big range of products made only by few states that make the measure of Complexity very inaccurate. From this observation, it is natural to use as the best measure of Complexity of products the ones $Q_p^W$ extracted from the Fitness algorithm applied to the trade of goods of all World countries, i.e. we take:

$$Q_p^W \equiv Q_p^B \equiv Q_p. \tag{3}$$

Indeed, the Complexities of the products obtained applying the Endogenous Fitness to the

World countries ($Q_p^W$) can be considered the same of the Complexities of the products inside Brazil ($Q_p^B$) and, therefore, we simply define them as $Q_p$.

Therefore, the algorithm consists of two steps:

1. We apply the (Endogenous) Fitness (Eq 2) to the World countries, as previously done in [5–7]. The criterion adopted to determine if a country $c$ is a "good" exporter of a given product $p$ is again based on the RCA extended to all World countries: we set $M_{cp} = 1$ if $RCA_{cp} \geq 1$ and $M_{cp} = 0$ otherwise (see the section *Database* for the source of the data). Applying the (Endogenous) Fitness algorithm to the matrix $M_{cp}$, after a sufficiently large number of iterations the algorithm converges to the fixed point so that, we obtain the respective Fitness $F_c$ for each country and the Complexity $Q_p^W$ for each product.

2. From the assumption Eq 3, we use as Complexity of the products exported by Brazilian states $Q_p$ the values obtained by the Fitness algorithm applied to the export of all World countries. Therefore, we use the information in the matrix $M_{sp}$ and the product Complexity $Q_p$ to calculate the Fitness of the Brazilian states through the following formula:

$$\begin{cases} \tilde{F}_s = \Sigma_p M_{sp} Q_p \\ F_s = \dfrac{\tilde{F}_s}{\langle \tilde{F}_s \rangle_s}. \end{cases} \qquad (4)$$

The relevance of developing the Exogenous Fitness measure is two folds. First of all, using world wide data we extract all the information to compute the Complexity of products to better compute the Fitness of states. Since the algorithm works by exploiting differences of capabilities, using world wide data we gain additional information related to the export baskets of countries with a wider range of Fitness and capabilities. Of course we still expect the two measures to be highly correlated in rank, in particular for a country like Brazil that contains such a vast array of development levels. As we will see in section *Comparison with other techniques*, this is indeed the case. The second reason is that the Exogenous Fitness allows to have for states Fitness values comparable with those of countries. Indeed, while the ranking between Exogenous and Endogenous Fitness are highly correlated, their actual values and distributions are vastly different. As detailed explained in the paper [33], while the ranking for the Fitness measure is always well defined, the shape of the matrix directly affects the convergence properties of the algorithm to a polarized distribution. Employing the Exogenous Fitness method we have smoothly changing values that allows for the forecasting exercises of section *Results*.

## Overview of Brazil

First, we analyze Brazil as a whole applying the (Endogenous) Fitness to World countries in the time interval from 1995 to 2015. In Fig 3 we show the matrix $M_{cp}$ of the World countries of year 2015 obtained by ordering the countries according to the Fitness and the products according to the Complexity. In that year, Brazil is ranked in the 44th/147 position (equivalent to the raw 103 in Fig 3).

In Fig 4, we show the dynamics of the World countries in the Fitness-Income plane emphasizing the BRIC countries (Brazil in green, Russia in blue, India in orange, and China in red). The figure shows that India and China have in 1995 lower values of GDP$_p$ than Brazil and Russia, but higher values of Fitness. According with [7], this difference justifies the dynamics in the plane of the four countries for the next years. Indeed, India and China continued their
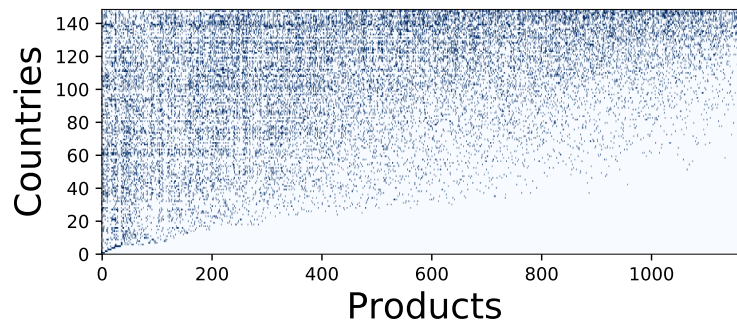
**Fig 3. The binary matrix $M_{cp}$ of the year 2015.** The rows of the matrix represent the World countries ordered according to their Fitness with row 0 for the country with the lowest Fitness and row 147 for the one with the highest Fitness. Analogously columns represent Products ordered in terms of their Complexity from the lowest one at column 0 to the highest one at column 1174. The elements $M_{cp} = 1$ are represented as blue dots.

https://doi.org/10.1371/journal.pone.0197616.g003

economic growth during the following years, while Russia and Brazil entered a period of recession [34].

In order to zoom on the differences among the dynamics of the BRIC countries, we analyze the variation of the Fitness of such countries during the interval from 2003 and 2013. The variation of the Fitness can have two different causes: (i) changes in the export basket, (ii) changes in the products Complexity. We can decompose the variation of Fitness [35] as:

$$\Delta \tilde{F}_c = \tilde{F}_c(t_1) - \tilde{F}_c(t_0) = \sum_p M_{cp}(t_1)Q_p(t_1) - \sum_p M_{cp}(t_0)Q_p(t_0) =$$

$$= \sum_p \Delta M_{cp} \frac{Q_p(t_1) + Q_p(t_0)}{2} + \sum_p \Delta Q_p \frac{M_{cp}(t_1) + M_{cp}(t_0)}{2}. \tag{5}$$
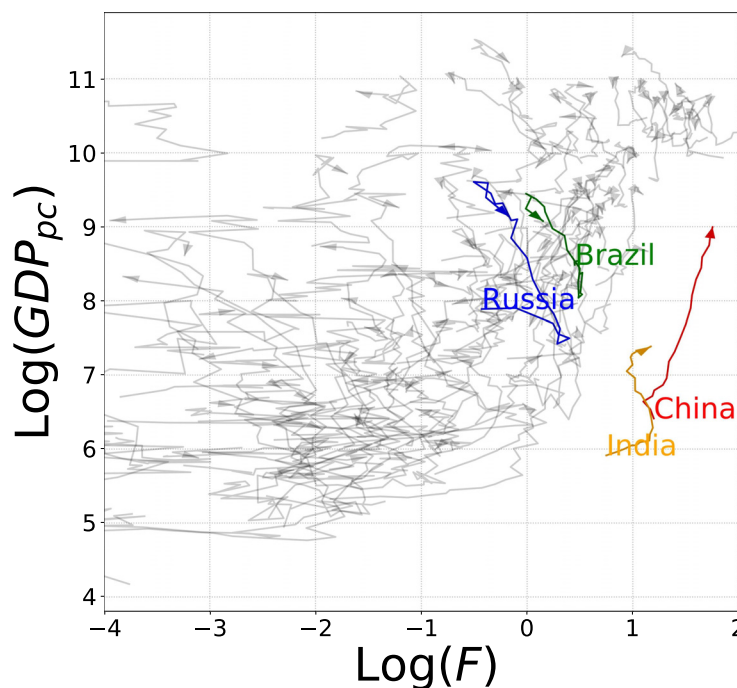


**Fig 4. Dynamics of the World countries in the Fitness-Income plane.** The figure shows the dynamics (from the year 1995 to the year 2015) of World countries in the Fitness-Income plane in logarithmic scale. We emphasize the BRIC countries: Brazil in green, Russia in blue, China in red, and India in orange.

https://doi.org/10.1371/journal.pone.0197616.g004

**Table 1. Fitness variation from 2003 to 2013 of BRIC countries.**

|  | Variation due to changes in the export basket | Variation due to changes in the products Complexities |
|---|---|---|
| Brazil | -43% | -6% |
| Russia | -37% | -21% |
| China | +32% | +18% |
| India | -18% | +2% |

where we have indicated with $\Delta X = X(t_1) - X(t_0)$ for a generic quantity $X$. The first term in the last step of the equation is the contribution to $\Delta \tilde{F}_c$ due to the variation in the export basket, while the second one is the term due to variation of products Complexities. In Table 1, we show both the percentage variations due to the two terms. The results show a deep decrease of both terms for Russia and we can see how the loss of competitiveness of Brazil is mostly due to the drop of products that were previously exported, and not so much related to the change in complexity of those products. In contrast China has increased its export basket and the Complexity of the exported products. Instead, India in 2013 exports more complex products, but has decreased its exports diversification.

Furthermore, we show in Fig 5 the products spectroscopy [32] for the BRIC countries of the year 2005 (dotted lines) and 2015 (filled colors). The figure shows that Brazil and Russia



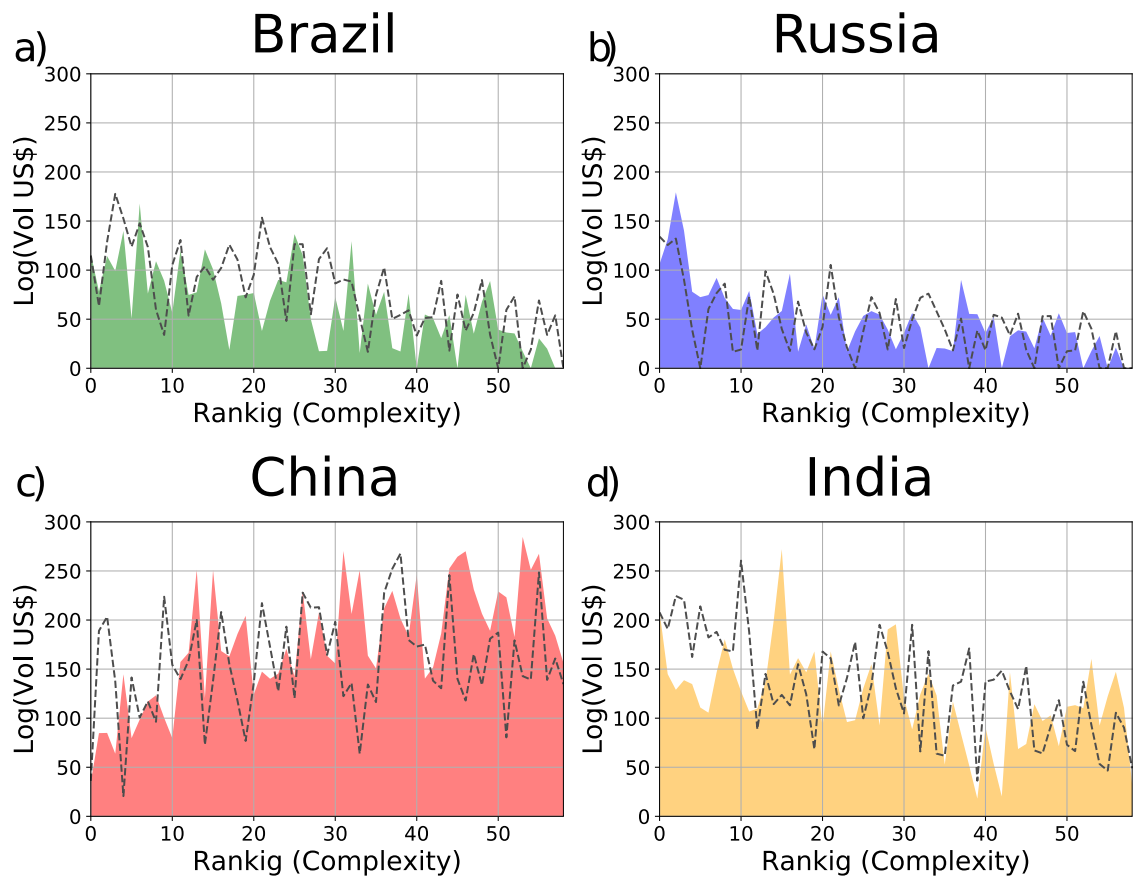**Fig 5. Products spectroscopy of the years 2005 (dotted lines) and 2015 (filled colors) of the countries: a) Brazil, b) Russia, c) China, and d) India.** The figures show the export volume (in US Dollars) of those states for each product with $M_{cp} = 1$ ordered in terms of their Complexity. The products have been grouped (10 for bin) and the export volumes of each product inside each bin have been summed.

have a high exportation only of simple products, while India and China have a high exportation of complex products.

Therefore, Figs 4 and 5 and Table 1, show that China and India both have a diversified export basket and export complex products. Such factors determine a high Fitness and consequently a growth of the GDP$_p$ in the subsequent years. On the contrary Brazil and Russia export simple products with a consequently low Fitness so that these countries entered a recession period [34].

In the next section we show the results of a deepened analysis of the internal economy of Brazil through the application of the Exogenous Fitness to the Brazilian states.

## Results

We applied the Exogenous Fitness algorithm to the Brazilian states in the time interval from 2000 to 2015 obtaining for each year both well-defined values of Fitness for each Brazilian state, and the ranking of states in terms of their Fitness (shown in Fig 6).
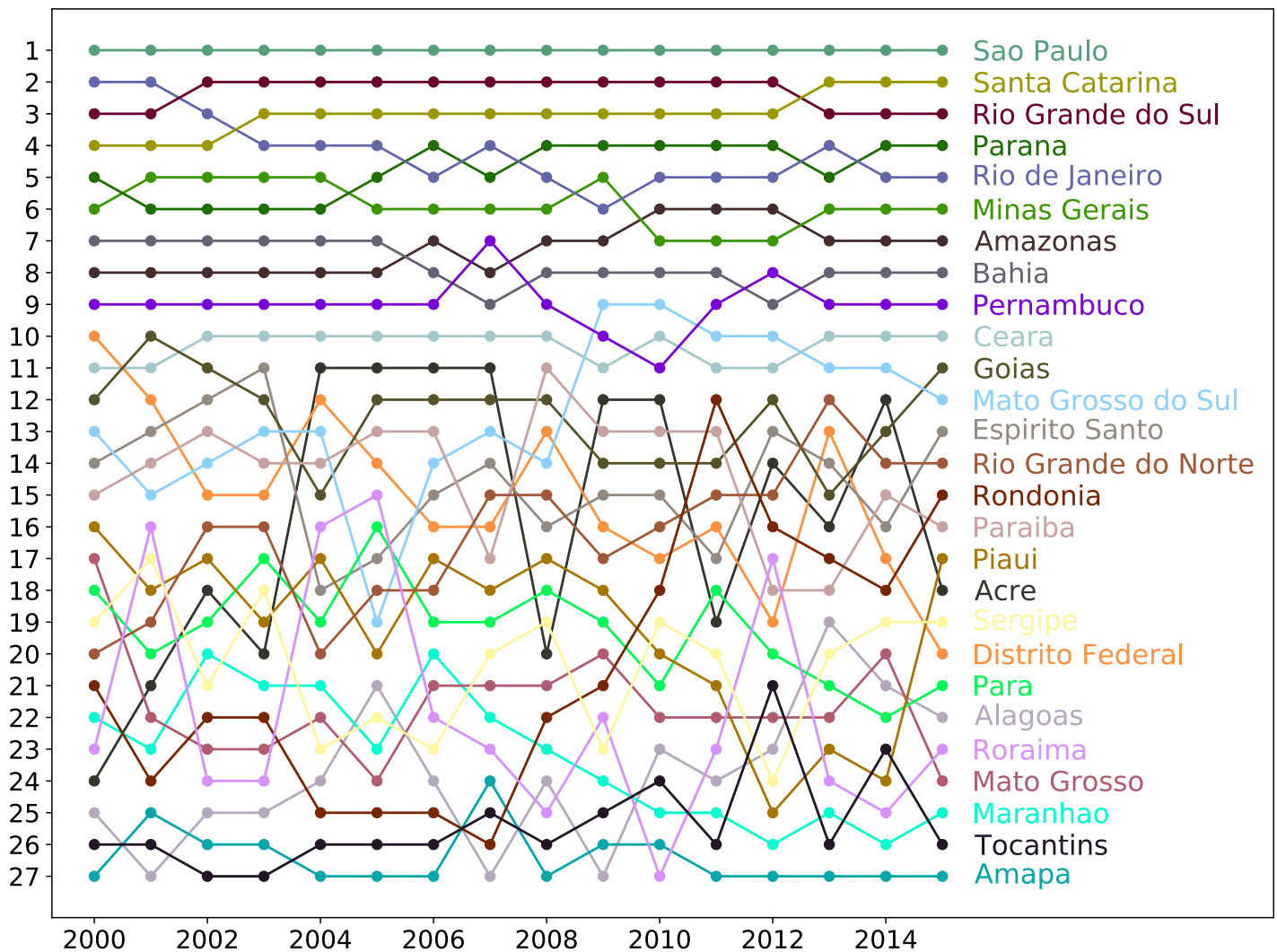


**Fig 6. Time evolution of the ranking of Brazilian states according to the Exogenous Fitness algorithm.** The figure shows the time evolution of the ranking of the Brazilian states according to the Fitness obtained through the Exogenous Fitness algorithm applied to the time interval 2000-2015.
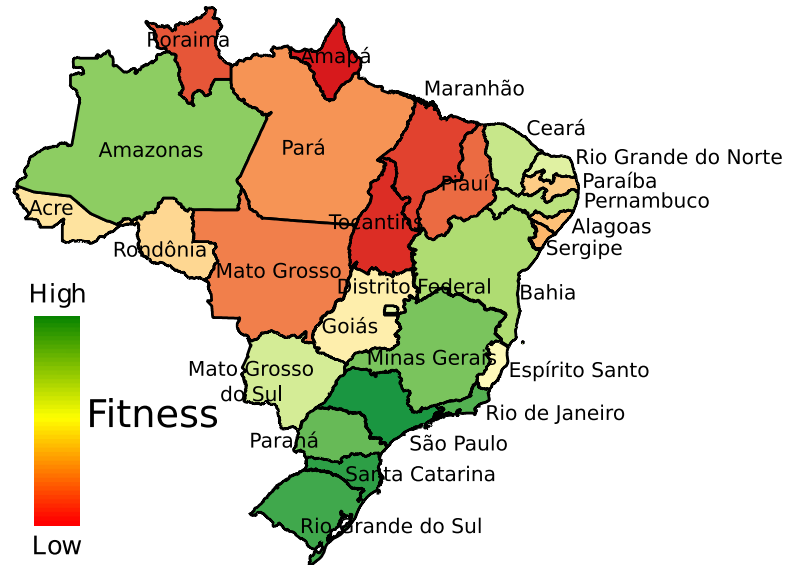
https://doi.org/10.1371/journal.pone.0197616.g006

**Fig 7. Fitness map of the Brazilian states.** The colors in the map vary from green (high Fitness) to red (low Fitness) and they show the differences of the Fitness among the Brazilian states.

https://doi.org/10.1371/journal.pone.0197616.g007

We show in Fig 7 a map of Brazil where each state is colored according to its Fitness. From the figure, it emerges Southern states have larger Fitness, and therefore have a better economic development, than Northern states. This result is in agreement with other monetary and non-monetary indices such as the Human Development Index (HDI) and the GDP [27].

Furthermore, we show in Table 2 the variation from 2003 to 2013 of the Fitness ($\Delta \tilde{F}_s$) for several states such as: São Paulo (1st in Fitness ranking of year 2013), Paraná (5th in Fitness ranking of year 2013), Ceará (10th in Fitness ranking of year 2013), and Roraima (24th in Fitness ranking of year 2013). From both Fig 2 and Table 2 we observe that São Paulo has a diversified export basket with high peaks in complex products and, at the same time, it increases both the export basket and the Complexity of the exported products in the considered time period. Paraná and Ceará, in contrast with the aggregate behavior of Brazil, in the same period grew in diversification becoming more competitive—even in the face of a minor decline in the complexity of their exported products. Roraima, on the contrarty, shows a deep decrease in the diversification.

As mentioned in the previous section, Fig 4 presents the dynamics of World countries in the Fitness-GDP$_P$ plane. It shows a high degree of heterogeneity of the dynamics of countries. Indeed, the plane can roughly be divided into two regions: one with an unpredictable "chaotic" regime of the evolution of countries, and the other with a predictable "laminar" regime. In order to overcome the limitations of linear regressions, Cristelli *et al* [7] proposed an

**Table 2. Fitness variation from 2003 to 2013 of the states: São Paulo, Paraná, Ceará, and Roraima.**

|  | Variation due to changes in the export basket | Variation due to changes in the products Complexities |
|---|---|---|
| São Paulo | +2% | +2% |
| Paraná | +59% | -7% |
| Ceará | +53% | -10% |
| Roraima | -37% | +6% |

https://doi.org/10.1371/journal.pone.0197616.t002

innovative data-driven non-parametric prediction scheme called the *Selective Predictability Scheme* (SPS). It is inspired by the so-called *method of analogues* [36, 37] and through a *measure of concentration* it delimits predictability regions inside the Fitness-Income plane. The measure of concentration consists in dividing the plane into a grid and analyzing the time evolution of the distribution of countries inside each box with at least five countries inside.

In analogy with what has just been explained for World countries, in Fig 8a, we show the time evolution of the real $GDP_p$ as a function of the Fitness (obtained implementing the Exogenous Fitness algorithm), for each Brazilian state in the period 2000-2015. The dotted black line in the figure shows the expected level of $GDP_p$ given the level of Fitness and it is the result of the minimization of the Euclidean distance of the states from the line, weighted by the state GDP. From the figure emerges an heterogeneous dynamics similar to the dynamics of World countries that cannot be analyzed through a linear regression. Also the *measure of concentration* is not appropriate in this case. Indeed the reduced number of Brazilian states (27) compared with the number of World countries (146) makes this measure inappropriate for the internal analysis of Brazil. In order to have a significant number of cells with at least five states, the granularity of the grid should be too broad to analyze the evolution of the distribution.

Therefore, in order to validate the predictability of the dynamics of the states in the Fitness-Income plane, here we develop a novel intuitive method, the *measure of direction*. First of all let us fix the time window $[t_1, t_2]$ in which we want to study the evolution of each state in the plane $\log(Fitness) - \log(GDP_p)$. The time lag $\Delta = t_2 - t_1$ has to be taken large enough to get a sufficient noise reduction in the dynamics. We choose $t_1 = 2003$ and $t_2 = 2013$. Second, we divide the plane in a fine grid of $100 \times 100$ cells and we define two bandwidth; one for the x-axis, and the other for the y-axis. For each cell, we define around its centroid a threshold area of sides given by the two bandwidths. Then, for each cell $k$ with at least three states at the time $t_1$ inside its threshold area, we computed the average dot product $\tilde{D}_k$:

$$\tilde{D}_k = \frac{2}{N(N-1)} \sum_{i<j}^{1,N} \hat{v}_i \cdot \hat{v}_j, \tag{6}$$

where $\hat{v}_i = \frac{\vec{v}_i}{v_i}$ where $\vec{v}_i = [log(F_i(t_2)) - log(F_i(t_1))]\hat{i} + [log(GDP_{p_i}(t_2)) - log(GDP_{p_i}(t_1))]\hat{j}$ and $\hat{i}$ and $\hat{j}$ are respectively the versors in the Fitness and $GDP_p$ directions. $N$ is the number of states with starting point inside the threshold area of cell $k$. The coefficient $\tilde{D}_k$ gives the average cosine among the versors of all states initially inside the threshold area of cell $k$ and varies from $(-1, 1]$. It measures the dispersion of the directions of evolution in the plane in the time window $[t_1, t_2]$ of all states initially in the threshold area of cell $k$: when it is close to 1 all states initially in the threshold area of cell evolve in a coherent parallel way. The smaller is $\hat{D}_k$ the larger the dispersion of these trajectories. A color map of the coefficient $\tilde{D}$ in the different cells is shown in Fig 8b. From the figure it emerges that there is a region where the directions of evolution of the states tend to be parallel (showed in green) and a region where the directions of motion tend to be unevenly directed (showed in red). Increasing/decreasing the bandwidths and, therefore, the threshold area only changes the resolution of the image, but the two regions remain well-defined. In Fig 8b we used an x-axis bandwidth 0.86, and a y-axis bandwidth 0.38, providing an almost continuous variation of the colors map.

In order to investigate which is the main direction of the versors in the green region and the further directions in the red region, we divided the plane into a broader grid (10x10). For each cell we sum all the vectors inside it and then we calculate the versor of the sum vector. We show the result in Fig 8c. From one hand, from the figure we can observe a region where the states tend to evolve in the same direction (shown in green). Therefore, in this region, the
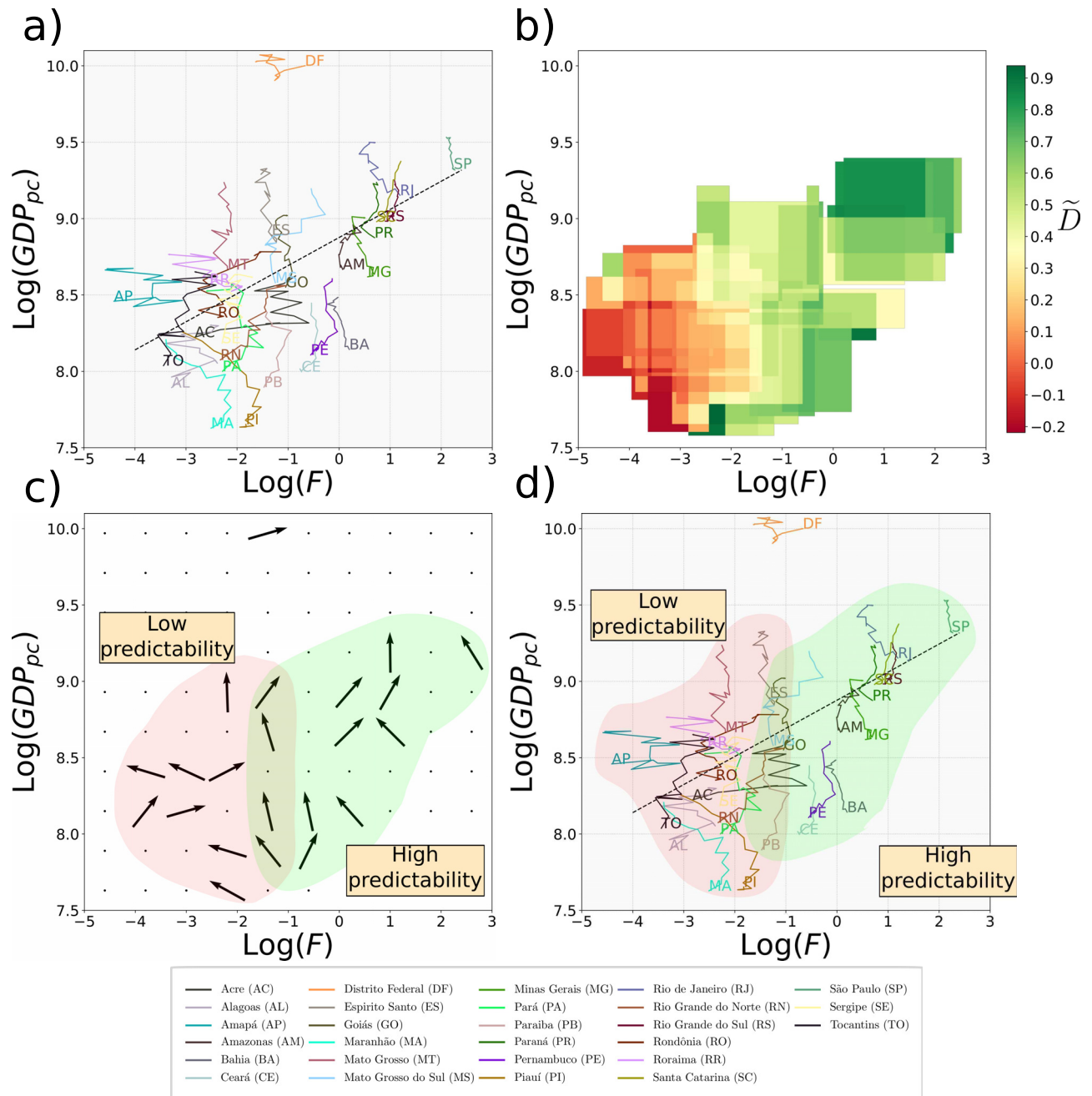
**Fig 8. Dynamics of Brazilian states in the Fitness-Income plane.** *a)* The figure shows the evolution (from 2000 to 2015) of the Brazilian states in the Fitness-Income plane in logarithmic scale. The dotted black line in the figure shows the expected level of $GDP_p$ given the level of Fitness and it is the result of the minimization of the Euclidean distance of the states from the line, weighted by the states GDP. *b)* The figure shows the coefficient $\widetilde{D}$ calculated considering a time window from 2003 to 2013. The color varies from green (where the versors of evolution tend to be parallel), to red (where the versors tend to be unevenly directed). *c)* The figure shows a grid where for each cell we calculate the versor of the sum vector. From the figure two regions appear: the first one where the versors tend to be parallel in the direction of a high $GDP_p$ (shown in green); and the second one where the versors tend to be unevenly directed (shown in red). Fig 8*b* and *c* together show that there is a region (green) of high predictability of motion in direction of a high $GDP_p$; and a region (red) of low predictability of motion. *d)* The figure shows the dynamics (from 2000 to 2015) of the Brazilian states in the Fitness-Income plane highlighting in green the states in the high predictability region and in red the states in the low predictability one.

https://doi.org/10.1371/journal.pone.0197616.g008

future evolution of countries is predictable with good confidence. On the other hand, another region (shown in red) can be detected where the versors tend to be unevenly directed. The dynamics of the states in this region is basically unpredictable. Furthermore, in the middle of the two, there is a region of transition, shown in the figure by the overlapping of the two colors.

Lastly, in Fig 8d we show the dynamics of the states in the Fitness-Income plane highlighting in green the states with high predictability of the motion and in red those with low predictability. From the figure emerges that states as Ceará, Pernambuco, and Bahia, despite having low values of GDP, are in a region of high Predictability and, therefore, they will probably continue to growth in the same direction. While for states as Acre, Tocantins, or Alagoas the dynamics is more chaotic and predictions are less reliable.

## Comparison with other techniques

In this section we compare the results obtained implementing the Exogenous Fitness with the results of the Endogenous Fitness and the ones published by Dataviva [27] obtained by applying the Economic Complexity Index (ECI).

### Exogenous Fitness and Endogenous Fitness

We apply the (Endogenous) Fitness algorithm to the Brazilian states in the time interval from 2000 to 2015 obtaining the time evolution of the ranking of the states according to such kind of Fitness (shown in Fig 9). Calculating the Spearman correlation coefficient between the ranking obtained through the Exogenous and the Endogenous Fitness for each year in the analyzed time interval, we obtain an average value $\tilde{\rho}_{ExEn} = 0.97$. This result shows a strong correlation between the rankings obtained through the two different Fitness algorithms.

The Endogenous Fitness algorithm provide us a well-defined annual ranking of the Brazilian states, but not well-defined quantitative values of Fitness and products Complexity. In fact, all Fitness values except one tend to zero. After a fairly high number of iterations, however, the ranking of states stabilizes, and there are no more changes of ranking among the states. This circumstance is already been studied [33] and it is due to the shape of the matrices $M_{sp}$. Indeed the *external area* (where $M_{sp} = 0$) is greater than the *internal area* (where almost all elements $M_{sp} = 1$) for each analyzed year.

### Exogenous Fitness and ECI

In Fig 10 we show the time evolution (from 2002 to 2015) of the ranking of the Brazilian states according to ECI, directly downloaded by the Dataviva platform [27]. Therefore, in order to compare the ranking obtained through the Exogenous Fitness algorithm and the ECI algorithm, we calculate the annual Spearman correlation coefficient between the two rankings in the period 2002-2015, obtaining an average value $\tilde{\rho}_{ExECI} = -0.14$. This result shows an almost total absence of correlations between the two rankings, i.e. between the two algorithms.

Indeed, already from a qualitative point of view, ECI ranking seems to be unrealistic. For example, it ranks rich states in GDP, but also with high HDI [27], such as Santa Catarina or Paraná, in the last positions (respectively 26th and 24th position in 2015). Moreover, the state of Alagoas (last in HDI ranking of 2014 [27]) is unrealistically ranked in 4th position in the 2015.

In Fig 11, we show the map of Brazil where each state is colored according to its ECI. From the figure, it emerges that there is no geographic coherence among the ECI of the different states. For instance the figure shows that the state of Santa Catarina has a high ECI, but it is in the middle between the states of Rio Grande do Sul and Paraná that have a low ECI.
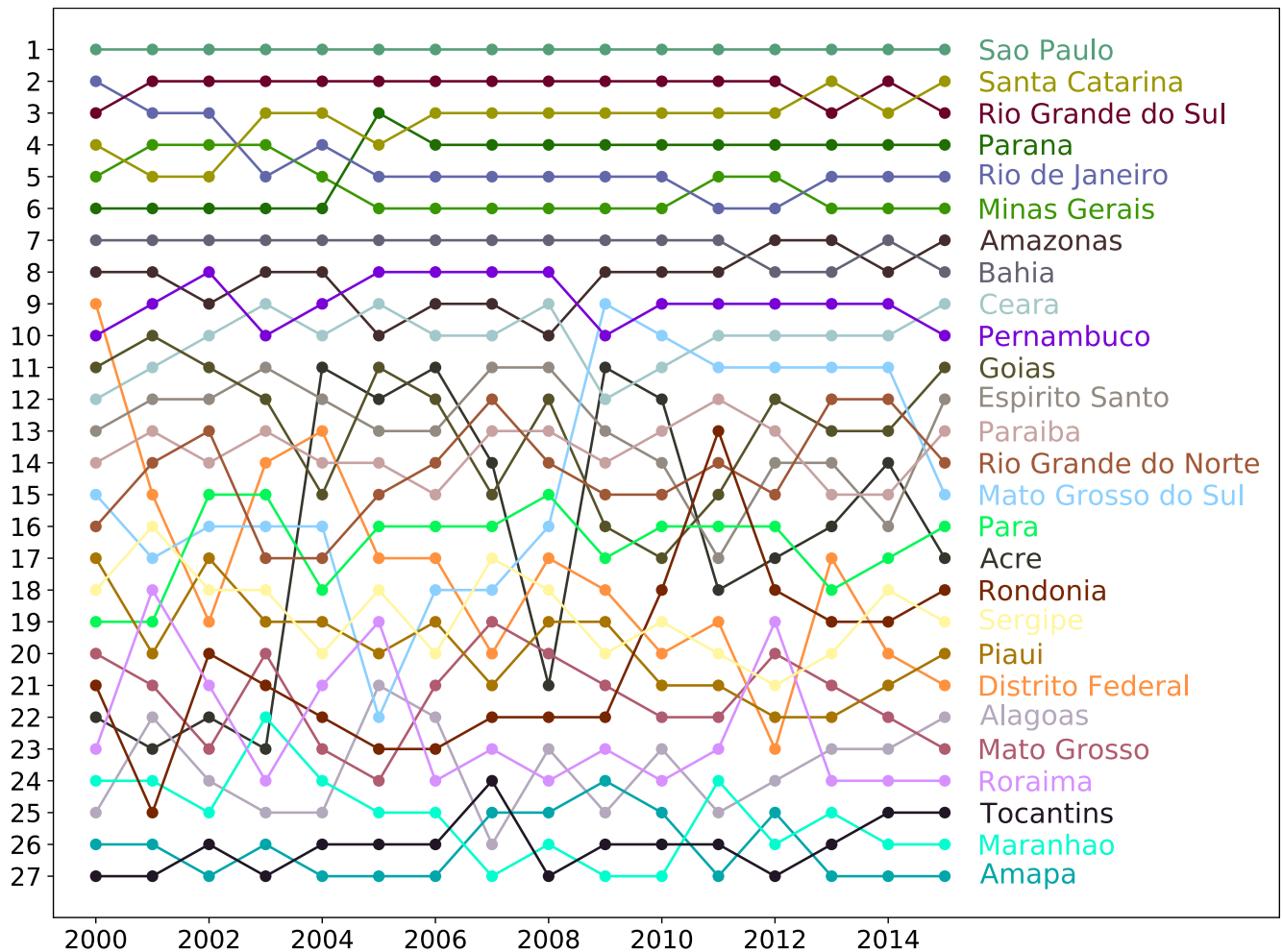
**Fig 9. Time evolution of the ranking of Brazilian states according to the (Endogenous) Fitness algorithm.** The figure shows the time evolution of the ranking of the Brazilian states in terms of the Fitness obtained through the (Endogenous) Fitness algorithm applied during the time interval 2000-2015.

Furthermore, we show in Fig 12a the evolution of Brazilian states in the ECI-Income plane, where the income is in logarithmic scale. In Fig 12b, we show the coefficient $\tilde{D}$ above defined but applied to ECI instead to $log(Fitness)$ and in Fig 12c the directions of motions. Differently from the results obtained through the application of the Exogenous Fitness (Fig 8), using the ECI index the dynamics of the states is unpredictable. Indeed, all the states except São Paulo and the Distrito Federal are concentrated in a small region of the plane and, therefore, totally indistinguishable.

## Discussion

In this paper we first compared the dynamics of Brazil in the Fitness-Income plane with the other BRIC countries. In Fig 4, we observed that IC (India and China) countries, both with a high Fitness compared to the BR (Brazil and Russia) countries, grow in GDP$_p$ for the entire analyzed time interval. Table 1 shows that IC improve the Complexity of export baskets in the analyzed time interval, and China even shows an improvement of the diversification. Instead, BR countries did not invested in diversification and in Complexity of the exported products
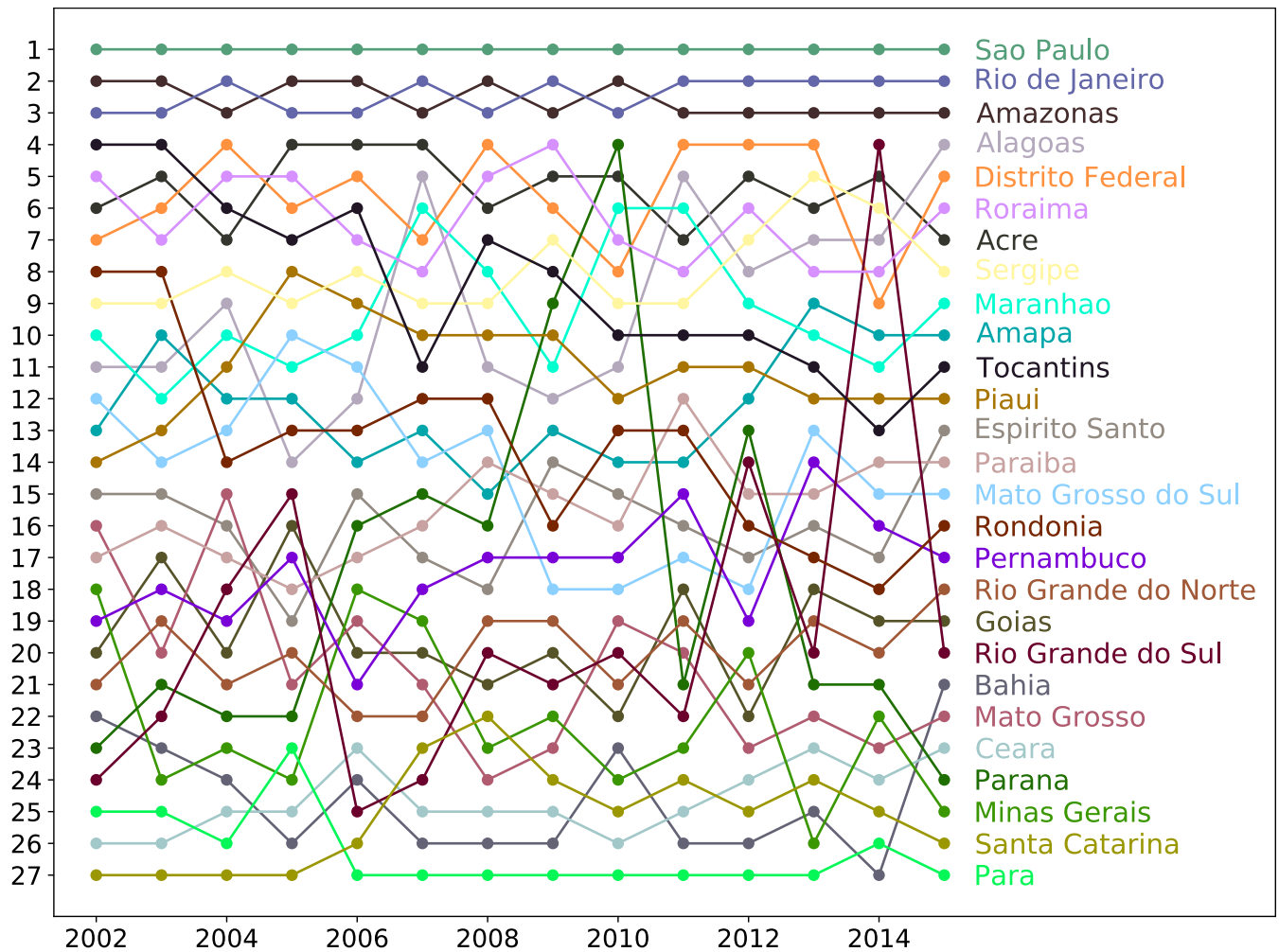
**Fig 10. Time evolution of the ranking of Brazilian states according to the ECI algorithm.** The figure shows the time evolution of the ranking of the Brazilian states during the period 2002-2015 in terms of the ECI, directly downloaded by the Dataviva platform [27].

https://doi.org/10.1371/journal.pone.0197616.g010

(as shown in Table 1). These results strengthen an hypothesis previously formulated in [7]: Fitness is the driving force behind growth.

In the second part of the paper, we introduced a new algorithm called "Exogenous Fitness" to calculate the Fitness of subnational entities and we applied it to the states of Brazil. The comparison between the Fitness and the GDP$_p$ showed an heterogeneous dynamics of the Brazilian states in the Fitness-Income plane. Indeed, two regions are distinguishable in the plane: one with high predictability and the other with low predictability. Here, we have shown that economic forecasting is possible for those states in the high predictability region, while it is not for those in the low predictability region. As a consequence of this analysis Fitness seems to be the driving force behind growth. Indeed, the dynamics in the high predictability region is characterized by high values of Fitness, while high value of GDP$_p$ is not a good signature of growth. The heterogeneous dynamics observed for the Brazilian states shows a strict analogy with the heterogeneous dynamics observed for the World countries [7]. Furthermore, by comparing the export "spectroscopy" of BRIC countries with the one of Brazilian states of São Paulo, Paraná, Ceará, and Roraima, and, comparing the variations of the Fitness, we observe that countries/states with diversified export baskets produce high complex products and grew in
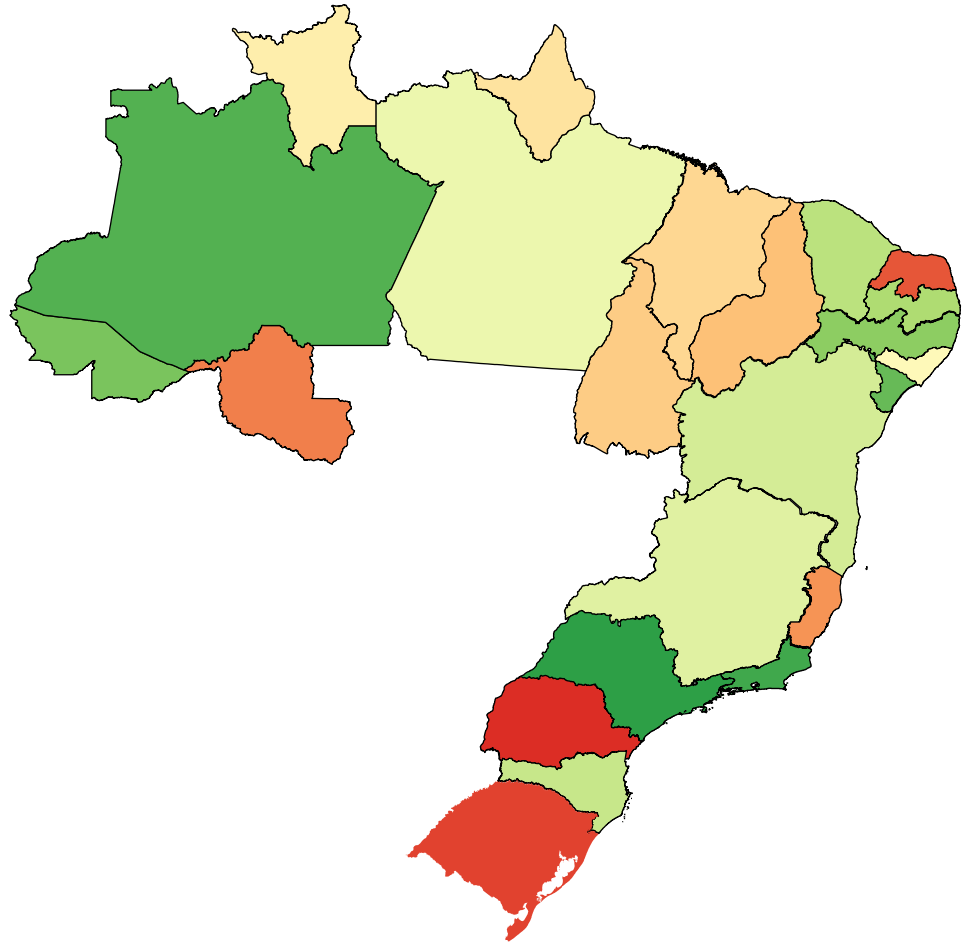
**Fig 11. ECI map of the Brazilian states.** The colors in the map vary from green (high ECI) to red (low ECI) and they show the variation of the ECI across the Brazilian states.

$GDP_p$ in the considered period. This observation can be important for the evaluation of perspectives of economic growth for Brazilian states, and, more generally, for developing countries.

The time evolution of the ranking obtained through the Exogenous Fitness algorithm shows that developed states in the top part of the ranking change little their positions, with a smooth slow motion. On the contrary states in the inferior part of the ranking changes drastically their position during the analyzed time-interval. These facts are probably due to the stability of the developed states that are in the high predictability region of the Fitness-$GDP_p$ plane and the instability of the states in the low predictability region.

Finally, we showed the non-correlation ($\tilde{\rho}_{ExECI} = -0.14$) between the ranking obtained though the Exogenous Fitness algorithm and the results of the ECI published by Dataviva [27]. Analyzing qualitatively the ranking of the states according to ECI, we argued that this ranking appears quite unrealistic. Therefore, we propose here the Exogenous Fitness algorithm as its valid substitute. Instead, comparing the Exogenous and (Endogenous) Fitness we obtained a strong correlation ($\tilde{\rho}_{ExEn} = 0.97$) for what concerns the ranking of states. This result shows that the two algorithmic tools are almost similar in identifying the ranking of the states, but
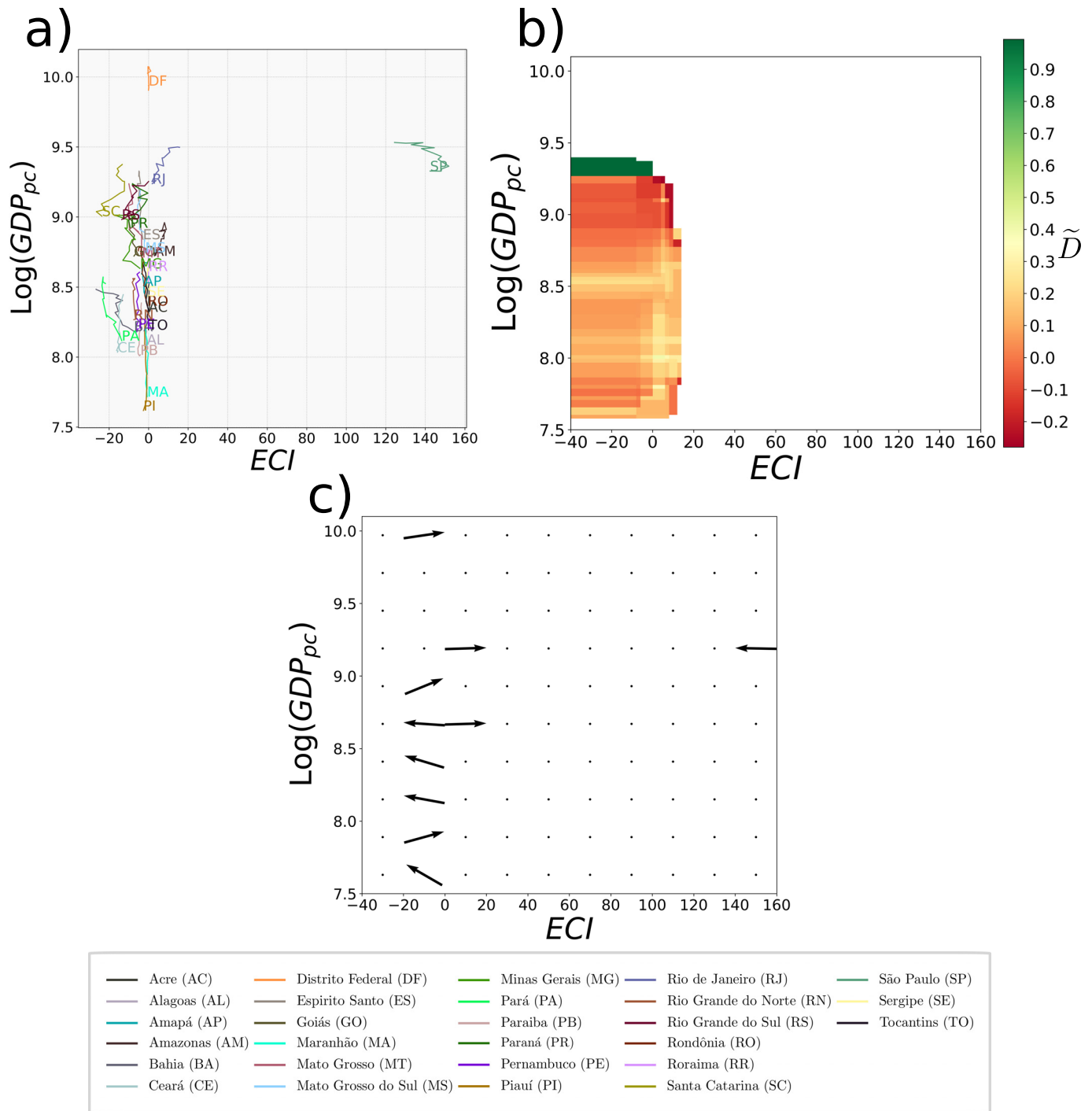
**Fig 12. Evolution of Brazilian states in the ECI-Income plane.** *a*) The figure shows the dynamics (from 2002 to 2015) of the Brazilian states in the ECI-Income plane, where the GDP$_p$ is in logarithmic scale. Only the state of São Paulo and the Distrito Federal appear to be clearly distinguishable from the rest of the states. All the others states are indeed concentrated in a small region of the graph. *b*) The figure shows the coefficient $\tilde{D}$ calculated considering the time interval 2003-2013. Colors vary from green (where the versors tend to be parallel), to red (where the versors tend to be unevenly directed). From the figure we can therefore verify that there is a low predictability of the evolution of all the states. *c*) Here we show a grid where for each cell we calculate the versor of the sum vector. From the figure we see that there is no privileged direction, indeed the vectors are unevenly directed.

https://doi.org/10.1371/journal.pone.0197616.g012

just the Exogenous Fitness algorithm provides also stable quantitative values of the Fitness, in addition to the ranking.

## Appendix A

### Dataset

The vast majority of data used in this paper is published by *DataViva* [27]. It is an open access platform that easily allows the access to a large amount of Brazilian socioeconomic data. The database is provided by the Brazilian *Ministries*: of *Employment* (MTPS), *Development, Industry and International Trade (MDIC)* and *Education* (MEC). The project is an initiative of the *Government of the state of Minas Gerais, Minas Gerais Investment, Trade Promotion Agency* (INDI) and the *Fundação de Amparo à Pesquisa do Estado de Minas Gerais* (FAPEMIG) [27, 38] in collaboration with the *Sistema Mineiro de Inovaçao* (SIMI) [39], *Big Data Corp* [40] and the *MIT Media Lab* [41]. The first version was published in November 2013 and the last one, the 3.0 version, in May 2015.

The platform includes data about imports/exports products, trade partners, occupation, economic activities, basic education, higher education and universities. All data are available in several levels of aggregation: region, state, mesoregion, microregion and municipality. The crossover among data and level of aggregations allows users to access more than 1 billion visualizations.

The visualization is made through some graph types, such as: Tree Map, Stacked, Geo Map, Network, Rings, Scatter, Compare, Occugrid, Line, Box Plot and Bar Chart. Furthermore, each data and aggregations is downloadable, and easily accessible through the API architecture [42].

Here, we use the export data of each Brazilian state for the entire time interval from 2000 to 2015. Furthermore, *DataViva* provides the data of total GDP and the total population for each state for the same time interval. Combining these with the GDP deflator $GDP_{defl}$ published by the *World Bank* [43], we find the real GDP *per capita* of each state as:

$$GDP_p^{real} = \frac{1}{N}\frac{GDP}{GDP_{defl}}100, \tag{7}$$

where $N$ is the total population of each state.

Concerning World export data, used to define the matrix $M_{cp}$ of the World countries and to calculate the products complexity, we use data from BACI dataset [29] that is grounded on the COMTRADE dataset [30]. The database, in its extension, contains data about more than 200 countries and 5000 products classified according to a 4 digit code with categorization *Harmonized System* 2007 [44]. Data are extracted from the year 2000 to 2015. The time evolution of the GDP *per capita* of each country is published by *World Bank* [45].

## Acknowledgments

## Author Contributions

**Conceptualization:** Felipe G. Operti, Emanuele Pugliese, José S. Andrade, Jr., Luciano Pietronero, Andrea Gabrielli.

**Data curation:** Felipe G. Operti.

**Formal analysis:** Emanuele Pugliese, Andrea Gabrielli.

**Investigation:** Felipe G. Operti, Emanuele Pugliese, Andrea Gabrielli.

**Methodology:** Felipe G. Operti, Emanuele Pugliese, Andrea Gabrielli.

**Software:** Felipe G. Operti.

**Supervision:** José S. Andrade, Jr., Luciano Pietronero, Andrea Gabrielli.

**Visualization:** Emanuele Pugliese.

**Writing – original draft:** Felipe G. Operti, Emanuele Pugliese, Andrea Gabrielli.

# References

1. Bureau of Economic Analysis. Regional Data. U.S. Department of Commerce. Retrieved 6th June 2017. Available from: https://www.bea.gov/.

2. Sawe B E. Indian States By GDP. WorldAtlas. Retrieved 5th November 2017. Available from: http://www.worldatlas.com/articles/indian-states-by-gdp.html.

3. Chepkemoi J. The Richest And Poorest States Of Brazil. WorldAtlas. Retrieved 5th November 2017. Available from: http://www.worldatlas.com/articles/the-richest-and-poorest-states-of-brazil.html.

4. Hidalgo C A and Hausmann R. The building blocks of economic complexity. PNAS. 2009. 106 no 26: 10570–10575. https://doi.org/10.1073/pnas.0900943106 PMID: 19549871

5. Tacchella A, Cristelli M, Caldarelli G, Gabrielli A, and Pietronero L. A New Metrics for Countries' Fitness and Products' Complexity. Scientific Reports. 2012. 2: 723. https://doi.org/10.1038/srep00723 PMID: 23056915

6. Cristelli M, Gabrielli A, Tacchella A, Caldarelli G, and Pietronero L. Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products. PLoS ONE. 2013. 8(8): e70726. https://doi.org/10.1371/journal.pone.0070726 PMID: 23940633

7. Cristelli M, Tacchella A, and Pietronero L. The Heterogeneous Dynamics of Economic Complexity. PLoS ONE. 2015; 10(2): e0117174. https://doi.org/10.1371/journal.pone.0117174 PMID: 25671312

8. Cristelli M, Ascagne C, Tacchella A, Cader M Z, Roster K I, and Pietronero L. On the predictability of growth. Policy Research working paper. World Bank Group, Washington, D C. 2017; WPS 8117. Available from: http://documents.worldbank.org/curated/en/632611498503242103/On-the-predictability-of-growth.

9. GDP ranking. The World Bank. Retrieved 23th June 2017. Available from: http://data.Worldbank.org/data-catalog/GDP-ranking-table.

10. Brazil Population. World Meters. Retrieved 23th June 2017. Available from: http://www.worldometers.info/world-population/brazil-population/.

11. Area Territorial Brasileira. Instituto Brasileiro de Geografia e Estatística-IBGE. Retrieved 23th June 2017. Available from: http://www.ibge.gov.br/home/geociencias/cartografia/default_territ_area.shtm.

12. Largest Countries in the World (by land area). World Meters. Retrieved 23th June 2017. Available from: http://www.worldometers.info/geography/largest-countries-in-the-world/.

13. Kuznets S. Economic Growth and Income Inequality. The American Economic Review. 1955. 45, 1: 1–28.

14. Furtado C. Formação Econômica Do Brasil. Rio de Janeiro: Fundo de Cultura. 1959.

15. Baer W, Kerstenetzky I, and Villela A V. The changing role of the State in the Brazilian economy. World Development. 1973. 1, 11: 23–44. https://doi.org/10.1016/0305-750X(73)90253-2

16. Sunkel O and Girvan C. Transnational Capitalism and National Disintegration in Latin America. Social and Economic Studies. 1973. 22: 1 132–76.

17. Hartmann D, Jara-Figueroa C, Guevara M, Simoes A, and Hidalgo C A. The structural constraints of income inequality in Latin America. arXiv. 2017.

18. Sen A. *Development as Freedom*. New York: ALFRED A. KNOPF INC. 1999.

19. Cajueiro D and Tabak BM. Ranking efficiency for emerging markets. Chaos, Solitons and Fractals. 2004. 22 349–352. https://doi.org/10.1016/j.chaos.2004.02.005

**20.** Chang E J, Lima E J A, and Tabak B M. Testing for predictability in emerging equity markets. Emerging Markets Review. 2004. 5 295–316. https://doi.org/10.1016/j.ememar.2004.03.005

**21.** Hidalgo C A, Klinger B, Barabási A-L, and Hausmann R. The product space conditions the development of nations. Science. 2007. 482–487. https://doi.org/10.1126/science.1144581 PMID: 17656717

**22.** Reynolds C, Agrawal M, Lee I, Zhan C, Li J, Taylor P, et al. A sub-national economic complexity analysis of Australia's states and territories. Regional Studies. 2017. 1–12.

**23.** Chávez J C, Mosqueda M, and Gómez-Zaldívar M. Economic Complexity and Regional Growth Performance: Evidence from the Mexican Economy. The Review of Regional Studies. 2017. 47, 2: 201–219.

**24.** Sbardella A, Pugliese E, and Pietronero L. Economic development and wage inequality: A complex system analysis. PLoS One. 2017. 19; 12 (9):e0182774. https://doi.org/10.1371/journal.pone.0182774 PMID: 28926577

**25.** Chotikapanich D, Griffiths W E, and Prasada Rao D S. Estimating and Combining National Income Distributions Using Limited Data. Journal of Business & Economic Statistics. 2007. 25, 1, 97–109. https://doi.org/10.1198/073500106000000224

**26.** O'Neill J. Building Better Global Economic BRICs. Goldman Sachs. Global Economics Paper. 2001. n 6.

**27.** DataViva. Retrieved 23th June 2017. Available from: http://www.dataviva.info/en/.

**28.** Balassa B. Trade Liberalisation and "Revealed" Comparative Advantage. The Manchester School. 1965. 33, 2: 99–123. https://doi.org/10.1111/j.1467-9957.1965.tb00050.x

**29.** Gaulier G and Zignago S. Baci: International trade database at the product-level. Retrieved 23th June 2017. Available from: http://www.cepii.fr/anglaisgraph/workpap/pdf/2010/wp2010-23.pdf.

**30.** UN COMTRADE database. Retrieved 23th June 2017. Available from: http://comtrade.un.org.

**31.** Battiston F, Cristelli M, Tacchella A, and Pietronero L. How metrics for economic complexity are affected by noise. Complexity Economics. 2014; 3: 1–22.

**32.** Cristelli M, Tacchella A, and Pietronero L. Economic Complexity. Measuring the Intangibles. A consumer's guide. Retrieved 23th June 2017. Available from http://www.lucianopietronero.it/wp-content/uploads/2016/02/economic_complexity_flyer-_v2.1_1.pdf

**33.** Pugliese E, Zaccaria A, and Pietronero L, On the convergence of the Fitness-Complexity algorithm, The European Physical Journal Special Topics. 2016. 225: 1893. https://doi.org/10.1140/epjst/e2015-50118-1

**34.** GDP Growth (Annual %). The World Bank. Retrieved 10th October 2017. Available from: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?end=2016&locations=BR-RU-CN-IN&start=2000.

**35.** Pugliese E, Chiarotti G L, Zaccaria A, and Pietronero L. Economic Complexity as a Determinant of the Industrialization of Countries: the Case of India. Retrieved 10th October 2017. Available from: http://pilhd.phys.uniroma1.it/PILgroup_Economic_Complexity/Publications_files/industrializationIndia_v4.pdf

**36.** Lorenz E N. Atmospheric predictability as revealed by naturally occurring analogues. Journal of the Atmospheric Sciences. 1969. 26: 636–646. https://doi.org/10.1175/1520-0469(1969)26%3C636:APARBN%3E2.0.CO;2

**37.** Lorenz E N. Three approaches to atmospheric predictability. Bulletin of the American Meteorological Society. 1969. 50: 345–349.

**38.** The Atlas of Economic Complexity. Center for International Development at Harvard University. Retrieved 23th June 2017. Available from: http://www.atlas.cid.harvard.edu.

**39.** Sistema Mineiro de Inovação: SIMI. Retrieved 23th June 2017. Available from: http://www.simi.org.br/.

**40.** Big Data Corp. Retrieved 23th June 2017. Available from: https://www.bigdatacorp.info/.

**41.** MIT Media Lab. Retrieved 23th June 2017. Available from: https://www.media.mit.edu/.

**42.** Api DataViva. Retrieved 23th June 2017. Available from: https://github.com/DataViva/dataviva-site/wiki.

**43.** GDP deflator. World Bank. Retrieved 23th June 2017. Available from: http://data.Worldbank.org/indicator/NY.GDP.DEFL.ZS?locations=BR.

**44.** Harmonized System. Retrieved 23th June 2017. Available from: http://www.wcoomd.org.

**45.** GDP per capita (current US dollar). World Bank. Retrieved 23th June 2017. Available from: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2016&start=1994.