Politecnico di Torino

Master of Science in Civil Engineering

Thesis

Economic Complexity: exploration of ranking metrics and investigation of causality between Complexity and Gross Domestic Product per capita.

Supervisors:

Prof. Francesco LAIO

Prof. Luca RIDOLFI

Ing. Carla SCIARRA

Candidate:

Luciano SARACENO

Academic year 2019/2020

Abstract

Several political economists (e.g. Ricardo, Smith) tried to explain the economic growth phenomenon without reaching a univocal conclusion. In recent years, a new discipline called Economic Complexity is emerging over classical economic theories, trying to give an answer to the fateful question "what is the secret of the wealth of nations?". According to Economic Complexity theories, it is possible to estimate the industrial competitiveness of a country just looking at its export basket, namely the products the country is able to export. In this approach, the trade data regarding exports is interpreted as a bipartite network in which countries are connected to the product they export. In order to classify countries and products belonging to this bipartite network, different algorithms have been developed, mutually correlating the complexity of countries to the one of products. At present, the commonly used Economic Complexity metrics are the Method of Reflections (MR) (Hidalgo & Hausman, 2009) and the Fitness and Complexity algorithm (FC) (Tacchella, Cristelli, Caldarelli, Gabrielli, & Pietronero, 2012). These two metrics diverge among them in the maths and in the results they supply. For this reason, a team from Politecnico di Torino have recently designed a new metric, called GENeralized Economic comPlexitY index (GENEPY) (Sciarra, Chiarotti, Ridolfi, & Laio, 2020) that is able to reconcile the two-existing metrics and furnishing a unique complexity ranking of countries and products. The first aim of this thesis is to analyse the three mentioned EC metrics, visualising the main differences and similarities among them. Particular attention has been paid to the complexity of products, because most of the works on Economic Complexity focused on the reliability of the results looking mainly at countries' competitiveness. Thus, through a detailed analysis of products' complexities and by exploring their evolution over time, we show that products have a crucial role on this complex system: the increase of a country's economic complexity translates into the presence of new complex products in its export basket; the presence of niche products in the network, especially if exported by developed countries, compromises the reliability of the results. The second aim is to investigate the causal relationship between the EC metrics and the most currently used economic competitiveness index, the Gross Domestic Product per capita (GDP). As first suggested by Sugihara et al. (Sugihara, et al., 2012), to carry this analysis we use the Convergent Cross Mapping technique. We show that a wide set of countries presents a strong causality correlation between GDP and EC metrics, suggesting the possibility to predict one from another. The prediction of GDP using only the EC algorithms represents the toughest challenge. However, our results on this subject show the possibility to consider Economic Complexity as a driving force for economic growth.

Contents

List of Figures ..... **iv**

List of Tables ..... **vi**

Acronyms ..... **vii**

1 Introduction **1**

1.1 Objectives and structure 4

2 The Economic Complexity **5**

2.1 Fundamentals of Networks Science 7

2.1.1 Centrality measures 9

2.1.2 Bipartite Networks 11

2.2 Economic Complexity metrics 14

2.2.1 The country-product matrix 14

2.2.2 Method of Reflections 15

2.2.3 Fitness and Complexity algorithm 18

2.2.4 Generalized Economic Complexity Index 20

3 Materials and Methods **25**

3.1 Materials 25

3.1.1 Input matrices 27

3.2 Detecting causation in Complex Systems 28

3.2.1 Convergent Cross Mapping algorithm 30

4 Results **35**

4.1 Complexity of countries 35

4.2 Complexity of products 44

4.2.1 Export basket composition 51

4.3 Causation between GDP and Complexity 54

5 Conclusions **63**

Bibliography **65**

Annex A1 ..... **69**

Annex A2 ..... **83**

List of Figures

Figure 2.1 : Economic Complexity Network. The bipartite network connecting countries and products is the result of the tripartite network in which countries are connected to their available capabilities and products are connected to the capabilities required to produce them (Gaulier & Zignago, 2009). 6
Figure 2.2 :The protein interaction map of yeast (Barabasi, 2016). 7
Figure 2.3 : Small network composed of nine nodes and seven links. 8
Figure 2.4: The two one mode projection of a bipartite network (Barabasi, 2016) 12
Figure 2.5: Graphical representation of Mcp for the year 2010. Reordering rows and columns by decreasing Diversity and Ubiquity it is evident the major triangular shape of the matrix (Cristelli, Gabrielli, Tacchella, Caldarelli, & Pietronero, 2013). 15
Figure 2.6: Graphical representation of diversity and ubiquity (Hausmann R. , Atlas of Economic Complexity, 2013) 16
Figure 2.7: Method of Reflections, year 2017. The Economic Complexity Index of countries corresponds the average of the complexity of the exported products. 18
Figure 2.8: Fitness and Complexity, year 2017. The Fitness of countries corresponds the sum of the complexity of the exported products. 20
Figure 2.9: Scatter plot of the second component of GENEPY, $X_{c,2}$, correlated to the ECI values over the countries squared degree - year 2017 (Sciarra, Chiarotti, Ridolfi, & Laio, 2020). 23
Figure 2.10: Scatter plot of the first component of GENEPY, $X_{c,1}$, correlated to the values of Fitness over the countries degree (year 2017) (Sciarra, Chiarotti, Ridolfi, & Laio, 2020). 24
Figure 3.1: Lorentz attractor and its coordinate time series projections (Sugihara, et al., 2012). 28
Figure 3.2: Shadow manifold of Lorentz attractor (Sugihara, et al., 2012) 29
Figure 3.3 : Convergent Cross Mapping based on the canonical Lorenz system in X,Y and Z (Sugihara, et al., 2012). 30
Figure 4.1: Ranking of countries, ordered by their ECI values, from 1995 to 2015. 37
Figure 4.2: Ranking of countries, ordered by their Fitness values, from 1995 to 2015. . 37
Figure 4.3: Ranking of countries, ordered by their GENEPY values, from 1995 to 2015. 38
Figure 4.4: Evolution of China's complexity from 1995 to 2015. Note that the values are normalized through the Frobenius' norm, in order to avoid excessive disparity between the results obtained with the different methods. 38
Figure 4.5: Scatter plot between Fitness of countries and ECI (year 2017). 39
Figure 4.6: Scatter plot between GENEPY and ECI (year 2017). 40
Figure 4.7: Scatter plot between GENEPY and Fitness of countries (year 2017) in logarithmic scale. 40

Figure 4.8: ECI of countries by changing input matrix (year 2017). 42
Figure 4.9: Fitness of countries by changing input matrix (year 2017). 43
Figure 4.10: GENEPY of countries by changing input matrix (year 2017). 43
Figure 4.11: Comparison of products' complexities. 45
Figure 4.12: Complexity vs Ubiquity (year 2017). 46
Figure 4.13: Top ten complex products (year 1995 – binary matrix). 47
Figure 4.14: Top ten products for year 1995 (RCA matrix). 49
Figure 4.15: Top ten products for year 1995 (USD matrix). 49
Figure 4.16: Top ten complex products for year 2014 (computed using binary matrix as input). 50
Figure 4.17: Export basket composition over years of DEU - Contribution of sectors Animals and Minerals to country's complexity using FC method. 52
Figure 4.18: Export basket composition of USA and NGA using GENEPY. 53
Figure 4.19: Evolution in time of Economic Indices. Each time series is normalized dividing each value by its Frobenius' norm. 55
Figure 4.20: Summary statistics of the CCM resulting from the reconstruction of the GDPpc from the GENEPY of China. The horizontal axe represents the length L and the vertical axe represents the CCM correlation. 56
Figure 4.21: China: CCM between GENEPY and GDPpc. The quantity ρ-GEN/M_{GDP} (blue line) refers to the correlation obtained by reconstructing GENEPY from GDP. The quantity ρ-GDP/M_{GEN} (red line) refers to the correlation obtained by reconstructing GDP from GENEPY. 57
Figure 4.22: Correlation obtained by reconstructing GDP at time t+lag using GENEPY at time t (CHN). 59
Figure 4.23: Scatter plot between GDP(t-4) and GENEPY(t) of CHN. Time series are normalized. 60
Figure 4.24: Cross Correlation between delayed time series of GDPpc and both first eigenvector (on the left) and second eigenvector (on the right). 61

List of Tables

Table 2.1 : Degree centrality scores concerning network in Figure 2.3......10
Table 2.2 : Eigenvector centrality scores concerning network in Figure 2.3......11
Table 2.3 : Degree centrality scores concerning network in Figure 2.4. 13
Table 2.4: Examples of the estimator function (Sciarra, Chiarotti, Laio, & Ridolfi, 2018). 21
Table 3.1: Products' Categories. 26
Table 4.1: Pearson correlation coefficient between Complexity of countries calculated by the three different EC metrics for the year 2017. 39
Table 4.2: Pearson correlation coefficient between Complexity of countries calculated by changing the input matrix. Results refer to year 2017. 42
Table 4.3: Pearson correlation coefficient between Complexity of products calculated by the three different EC metrics for the year 2017. 44
Table 4.4: China's time series of Complexity and GDP per capita. 54
Table 4.5: CCM correlations between GENEPY and GDP per capita of China......57
Table 4.6: CCM between GDPpc (t+lag) and GENEPY(t) of China......59

Acronyms

CCM	Convergent Cross Mapping
EC	Economic Complexity
ECI	Economic Complexity Index
FC	Fitness and Complexity algorithm
GDP	Gross Domestic Product
GDPpc	Gross Domestic Product per capita
GENEPY	Generalized Economic Complexity Index
MR	Method of Reflections
PCI	Product Complexity Index
RCA	Revealed Comparative Advantage
SVD	Singular Value Decomposition
TSS	Total Sum of Squares
UC	Unique Contribution
USD	United States Dollars

CHAPTER 1

1 Introduction

Since the early 1980s, China have been characterized by an astounding economic growth, accompanied by profound institutional changes and economic reforms (Kuhn, 2013). The country's Gross Domestic Product (GDP) grown on average by 10% in the last 30-40 years, transforming China from a backward agricultural economy to a global economic power (International Monetary Fund, n.d.). A considerable number of empirical and theoretical studies have been conducted to investigate what is the source of China's economic growth at both national and provincial levels, and still debating about the driving-force of this unsettling development (Liang & Teng, 2005). The search for the secret of economic development and industrial sophistication is a difficult challenge also outside China's boundaries, because the economic growth phenomenon is extremely complex, involving different factors: institutions, laws, education, infrastructures, corruption, natural resources availability, pollution and more. Nowadays all these *Big Data* (English term, typical of statistics and information technology, defining a particularly extensive set of data in terms of volume and variety) are almost totally available, but it still seems utopic to develop methods and algorithms able to estimate the economic competitiveness of a country taking into account all of these variables. In recent years, a new discipline, called *Economic Complexity*, is making its way through classical economic theories, trying to intensify the level of scientific contents in economic models (Pietronero, Cristelli, & Tacchella, 2013) and, introducing reasonable simplifying hypotheses, quantify economic competitiveness of countries by using a data-driven approach.

By furnishing strong theoretical tools and encouraging innovative steps finalized to a better understanding and synthesis of different empirical realities from increasingly available datasets; both networks Physics and Complex Systems theory are contributing to our understanding of social, ecological, environmental, and economic processes (Tu, Carr, &

Suweis, 2016). Complexity is a fundamental characteristic of our world (Bar-Yam, 2014), thus characterized by a large number of components which may interact with each other and require development of new methods and algorithms. It has been shown (William & Martinez, 2000) that remarkably simple models, employing only few empirical parameters, are able to successfully predict the fundamental structural properties of the most complex food chains and ecosystems in nature. Following this philosophy, in 2009 (Hidalgo & Hausman, 2009) Cesar A. Hidalgo and Ricardo Hausmann proposed to estimate the industrial competitiveness of a country, merely looking at the products a country is able to export, hence assuming that a product embeds the effects of all the others variables on economic development. In their view, modern societies are the result of the accumulation of '*productive knowledge*' (the knowledge required to make a product) developed by the humankind during the past two centuries (Hausmann R. , Atlas of Economic Complexity, 2013). Our lifestyles have been made easier because of the advent of large number of innovative stuff: think about cars, smartphones, TVs, microwaves, medicines, vaccines and many others. Therefore, we should think of a product not as a physical object but as the set of all the industrial knowledge necessary to produce it. As an example, the true value of a computer is that it embeds knowledge about electronic engineering, mechanical engineering, magnetism, optics, data processing systems, programming and all the necessary skills for hardware functioning and software development. Markets play a crucial role in this process, since by allowing us to buy products, and thus productive knowledge from people from all over the world. Each individual can store a limited amount of productive knowledge, entailing that the amount of knowledge possessed by a society depends on how much its individuals' knowledge is diversified and how more complex the network of interaction between people is, allowing to combine different knowledge into more sophisticated ones.

Adam Smith, father of classical economy, already believed that the wealth of nations resides in the diversification of labour (Smith, 1776). Diversification is a central element in economy and revokes concepts characteristic in ecology, like the adaptability of the species to an evolving environment (Pietronero, Cristelli, & Tacchella, 2013); suggesting that economic efficiency increases with specialization of people and firms in different fields and implying that development is linked to the growing number of individual activities and thus to the complexity derived from the interaction between them (Hausmann R. , Atlas of Economic Complexity, 2013).

We use the term *capabilities* to indicate knowledge embedded in each person and permitting them to perform a certain task. We can therefore deduce that the more capabilities a country possesses, the more complex it is, reflecting high levels of economic and industrial

sophistication. Unfortunately, capabilities cannot be directly measured, but we can use the products exported by a country in order to indirectly measure them. According to Hidalgo and Hausmann we can indirectly measure the capabilities available in a country by considering each capability as a building block, or Lego piece (Hidalgo & Hausman, 2009). In this analogy, we can think of a product as a Lego model and a country as a bucket of Legos: countries will be able to make products (Lego models) as long as they possess all the necessary capabilities (Lego pieces).

The *Economic Complexity* is a measure of the quality and the quantity of capabilities a country possesses and thus a measure of how much intricate the network of interaction between its knowledge owners is. The aim of the Economic Complexity theories is to estimate the productive knowledge knowing only the products a country is able to export, because products itself embeds capabilities required to be produced. Naturally, one can affirm that countries may be able to make goods that they do not export. However, since they are not able to introduce such products into the global market is an indicator of their low complexity. Another observation is that trade-data embeds only goods and not services. This one is perhaps the most restrictive limitation because in the last years the trade of services has become increasingly important, but it is still difficult to track in a trustworthy way (Hausmann R. , Atlas of Economic Complexity, 2013) and therefore unreasonable to include it into the Economic Complexity estimation process. Different algorithms, mutually correlating the complexity of countries to the complexity of products, have been developed in order to classify the industrial competitiveness of countries and products. At present, the commonly used Economic Complexity metrics are the Method of Reflections (MR) (Hidalgo & Hausman, 2009) and the Fitness and Complexity algorithm (FC) (Tacchella, Cristelli, Caldarelli, Gabrielli, & Pietronero, 2012). These two metrics diverge among them in the maths and in the results they supply. For this reason, a team from Politecnico di Torino have recently designed a new metric, called GENeralized Economic comPlexitY index (GENEPY) (Sciarra, Chiarotti, Ridolfi, & Laio, 2020) that is able to reconcile the two-existing metrics, furnishing a unique complexity ranking of countries and products.

1.1 Objectives and structure

The goals of this thesis are mainly two.

First, we would like to contribute with our work to the understanding of how the different Economic Complexity metrics work, visualising the main differences and similarities among them and focusing on the evolution over time of the complexity of both countries and products; particular attention has been paid to the complexity of products, because most of the works on Economic Complexity focused on the reliability of the results looking mainly at the countries' competitiveness.

The second one is the investigation of the causal relationship between the Economic Complexity metrics and the most currently used economic competitiveness index, the Gross Domestic Product per capita.

The report is organized as follows:

Chapter 2*: The Economic Complexity,* chapter dedicated to an in-depth description of available and developing metrics of economic complexity.

Chapter 3*: Materials and Methods,* chapter dedicated to the description of the input data and the explanation of the Convergent Cross Mapping technique used to investigate causation between EC and GDP.

Chapter 4*: Results,* chapter dedicated to the discussion and visualization of the main difference and similarities among the Economic Complexity metrics, focusing on the evolution of complexity over time of both countries and products, the response of the different algorithms at different inputs, and the correlations between Economic Complexity indices and Gross Domestic Product per capita.

Chapter 5: *Conclusions,* last chapter dedicated to highlighting the accuracy of Economic Complexity metrics in describing economic status of some countries and to reflecting on the possibility to consider Economic Complexity as a driving force for economic growth.

CHAPTER 2

2 The Economic Complexity

The idea that nations' prosperity depends on their ability to develop increasingly complex and innovative products, able to conquer world markets, is as old as the first economic theories born after the Industrial Revolution. However, complexity is a difficult attribute to measure, and it is often assigned a priori (for example, it is common to say that a computer is more complex than an apple).

The economic growth phenomenon is rather difficult to analyse because of its complex and disaggregate nature. Even so, recent availability of Big Data, techniques and ideas from the Science of Complex Systems allow us to consider the interaction between economic components, introducing new methods and algorithms able to characterize the economic structure of the world. In the last years, a new discipline, called Economic Complexity, is making its way through classical economic theories, trying to intensify the level of scientific contents in economic models (Pietronero, Cristelli, & Tacchella, 2013) and quantify complexity of both products and countries by using a data-driven approach.

The aim of the Economic Complexity theories is to estimate the productive knowledge of countries knowing only the products a country is able to export. A country can export a product only if it has all the raw materials and industrial know-how to produce it. Specifically, we use the term capabilities to indicate that set of human capital, physical capital, laws, institutions, and more needed to produce a given product (Abdon, Bacate, Felipe, & Kumar, 2010). Obviously, each product requires a different number of capabilities to be produced and each country possesses different capabilities (economically developed countries probably possess many of them). Thus, capabilities can be used to quantify complexity of both countries and products. Countries will be able to make all the products for which they possess all the necessary capabilities. As a direct consequence it is possible to estimate the capabilities available in a country by just looking at the products the country is able to make. This is

equivalent to consider a tripartite network in which countries are connected to their available capabilities and products are connected to the capabilities required to produce them (*Figure 2.1*). The result of this tripartite network is a bipartite network in which countries are connected to the product they export (*Figure 2.1*), allowing to look at country-product associations just by employing international trade data.

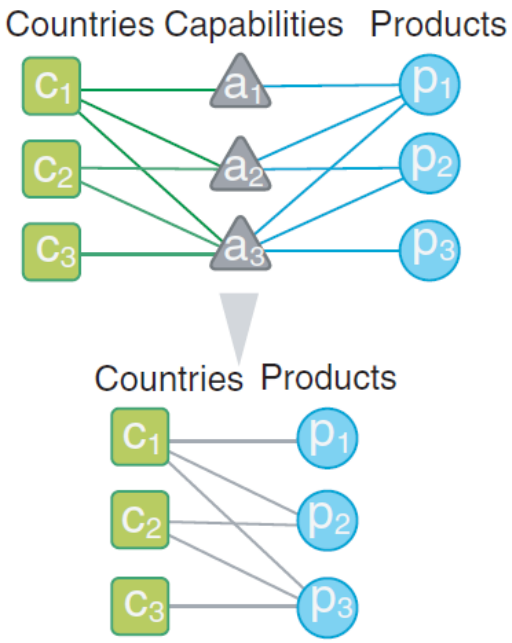


Figure 2.1 *: Economic Complexity Network. The bipartite network connecting countries and products is the result of the tripartite network in which countries are connected to their available capabilities and products are connected to the capabilities required to produce them* **(Gaulier & Zignago, 2009)**.

Quantifying the economic complexity of countries and products is therefore equivalent to characterising the structure of a bipartite network in which each country is linked to the products in its export basket. The network characterization process consists in the ranking of the nodes composing it (countries and products) according to the more important they are. Different algorithms have been developed to classify the industrial competitiveness of countries and products, mutually correlating complexity of countries to complexity of products. In order to understand how these algorithms work, and thus what it means to be important in a network, it is necessary to recall some fundamental concepts from Network Science.

2.1 Fundamentals of Networks Science

A network is, in its simplest definition, a set of points connected to each other by lines (Newman, Networks: An introduction, 2010). In order to understand how a complex system works, it is often useful to represent it as a network whose nodes represent the system's components and links their interactions (Barabasi, 2016). Network theory finds application in many systems of scientific interest, examples include the Internet, the human society, the protein interaction (*Figure 2.2*), the neural networks, the mobile-phone calls and more.

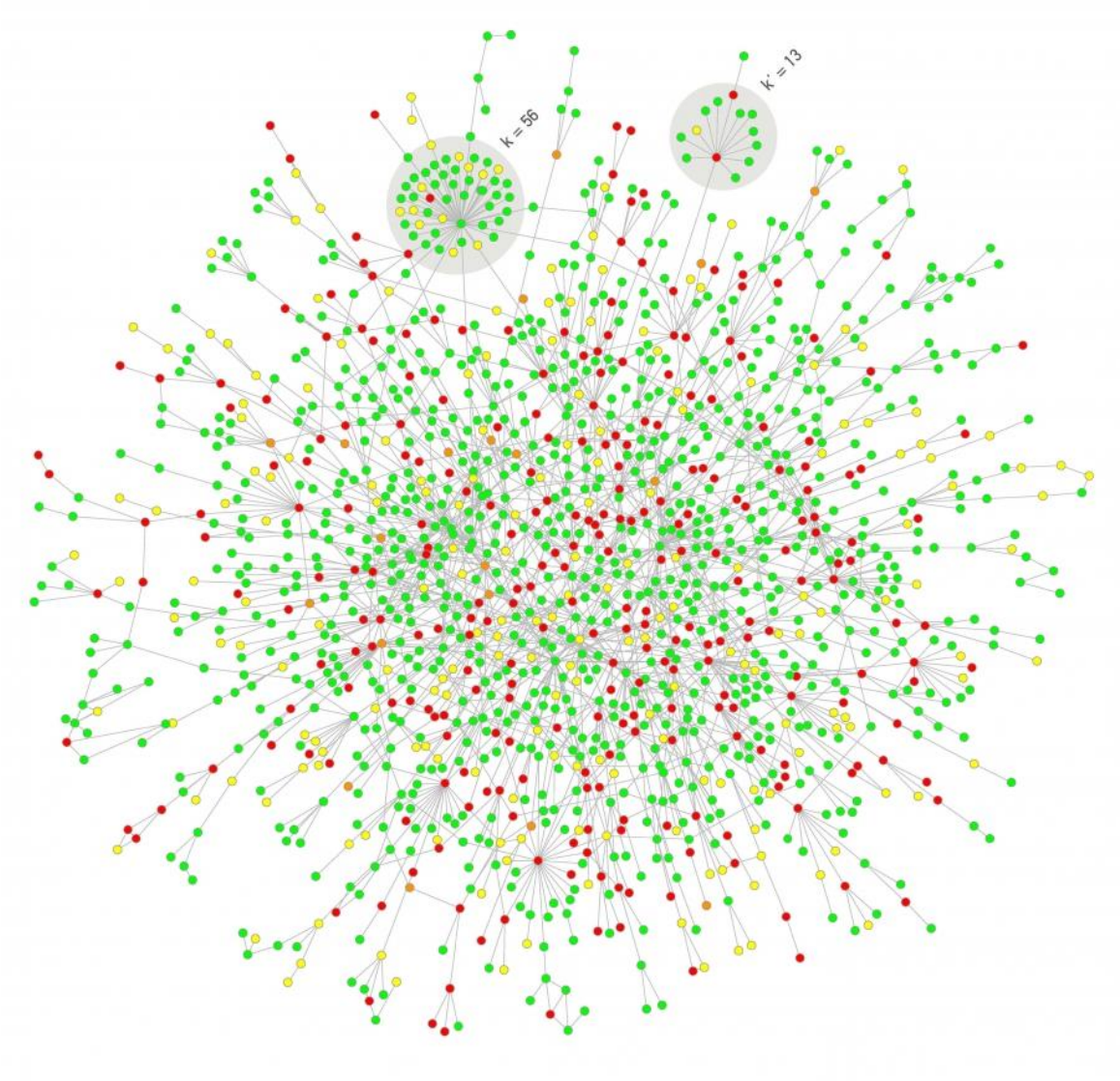


Figure 2.2 *:The protein interaction map of yeast* **(Barabasi, 2016)**.

Thanks to network science it is possible to simplify a complex real system with an abstract structure preserving only the essential characteristics that define interactions between the parts. Nodes and links can carry additional information to capture more details, but when a system turns to a network representation usually we lose a lot of information (Newman, Networks: An introduction, 2010).

A *directed network* is a network in which links connecting the nodes have a direction, pointing from one node to another one. An example of directed network is the food chain web, in which energy flows from prey to predator. Instead, an *undirected network* is a network in which links have no direction.

Mathematically, the information carried by a network can be stored in the so-called *adjacency matrix*. As an example, consider an undirected network consisting of nine nodes and seven links (*Figure 1.2*): its adjacency matrix, A, is a 7x7 matrix. The elements of the matrix, Aij, are equal to 1 if between node i and node j a link exists, otherwise Aij is equal to 0.

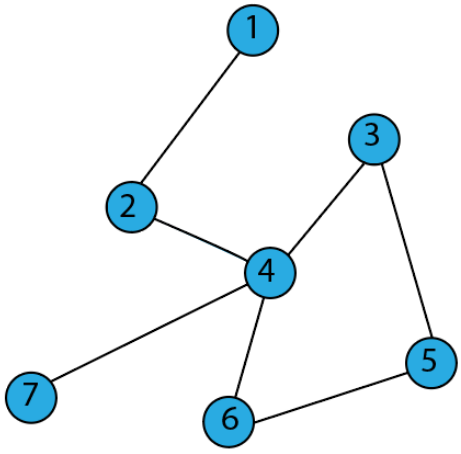


***Figure 2.3** : Small network composed of nine nodes and seven links.*

The adjacency matrix representing the network illustrated in *Figure 2.3* have the following form:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

This is just a simple representation, but there are cases in which the links are not all the same. We define this type of networks as weighted networks and the adjacency matrix describing them is characterized by elements different from 1. In addition, all the elements belonging to the main diagonal are equal to zero because in the network there are no self-edges (a self-edge is a link between a node and itself).

2.1.1 Centrality measures

The characterization of a network's structure is not a simple operation, especially if the network consists of millions of nodes and it is therefore extremely difficult to visualize. *Centrality measures* are some of the essential tools used in network topology characterizations, answering to the question "Which are the most important nodes in a network?" (Newman, The mathematics of networks, 2006).

Different techniques have been developed in order to state the importance of a node in a network, focusing on different concepts and definitions of the meaning of centrality. In this section we are going to see just the two simplest measures of centrality: the *degree centrality* and the *eigenvector centrality*.

2.1.1.1 Degree Centrality

Both historically and conceptually, the first measure of centrality has been the *degree centrality*. The *degree* of a node is the number of links connected to it. If the network is directed, we define two different measures of degree centrality, called *indegree* and *outdegree*. Let A be the adjacency matrix describing an undirected unweighted network and let A_{ij} be the elements of this matrix; we define the degree centrality of the i-th node as:

$$k_i = \sum_{j=1}^{n} A_{ij}.$$

For instance, the degree centrality scores of the nodes belonging to the network illustrated in *Figure 2.3* are:

node	degree centrality score
1	1
2	2
3	2
4	4
5	2
6	2
7	1

Table 2.1 *: Degree centrality scores concerning network in Figure 2.3.*

Examples where it is useful to use the degree centrality are the social networks, since it is reasonable to state that individuals that have more connections are in a more influential position in the network (Newman, Networks: An introduction, 2010). However, degree centrality is just a simple measure, but it still represents an excellent preliminary analysis tool.

2.1.1.2 Eigenvector Centrality

A natural evolution of the degree centrality is the *eigenvector centrality*. Where the *degree centrality* is limited to numbering the connections of a node, *eigenvector centrality* takes into account the concept that not all connections in a network have the same weight during the topology characterization process. Using the social networks example again, it is reasonable to argue that, in general, connections with people who are themselves influential will make a person more relevant than connections with less influential people (Newman, The mathematics of networks, 2006). Thus, connections to highly connected nodes contribute more to the centrality score, of a selected node, than connections to lowly connected nodes (Negre, et al., 2018). Eigenvector centrality assigns each node a score that is proportional to the average of the scores of its neighbours (Newman, Networks: An introduction, 2010). Let A be the adjacency matrix describing an undirect unweighted network and let A_{ij} be the elements of this matrix, we define the eigenvector centrality of the i-th node as:

$$x_i = \frac{1}{\lambda}\sum_j A_{ij}x_j$$

Where λ is a constant and j are neighbours of node i. Defining the vector of centralities as $x = (x_1, x_2, x_3...)$, we can equivalently rewrite the equation in matrix form as:

$$\lambda x = Ax$$

Now it is clear that x is an eigenvector of A and λ is its associated eigenvalue. Using the *Perron-Frobenius theorem* we can also state that λ must be the largest eigenvalue and x its corresponding eigenvector (Newman, The mathematics of networks, 2006).

For instance, the eigenvector centrality scores of the nodes belonging to the network illustrated in *Figure 2.3* are:

node	eigenvector centrality score
1	0.055
2	0.128
3	0.165
4	0.242
5	0.142
6	0.165
7	0.104

***Table 2.2** : Eigenvector centrality scores concerning network in Figure 2.3.*

Looking at *Table 2.2* it can be observed that the eigenvector centrality score of node 7 (0.104) is higher than node 1 (0.055) even if they have the same number of links; the difference occurs because node 7 is connected to node 4 (the most central node in the network) and then its link weights more than the one of node 1, connected instead to node 2.

Hence, eigenvector centrality provides a measure of centrality that is the result of both the number and the quality of links possessed by a node. A variant of this centrality metric is *Google's page rank,* developed by the well-known web services company.

2.1.2 Bipartite Networks

A bipartite network is a network whose nodes can be divided into two disjointed sets. In such a network one set represents the original nodes and the other one represents the group to which they belong. Therefore, there are no links between vertices belonging to the same set, but only between nodes of different ones. The Economic Complexity network is a perfect example of bipartite network, in which products (original nodes) are connected to countries (groups) they belong.

Considering for example a bipartite network whose nodes can be divided in two disjoint sets U and V: for each set we can make a one mode projection, that is another network in which two nodes of the same set are connected if in the bipartite network both have a link with the

same node belonging to the other set. A graphical representation of this concept is given in *Figure 2.4*.

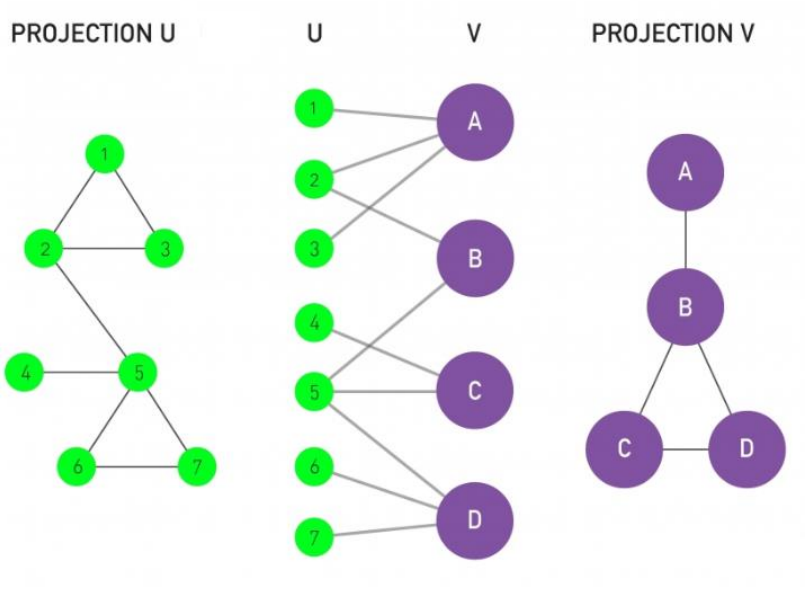


***Figure 2.4**: The two one mode projection of a bipartite network* **(Barabasi, 2016)**

As a consequence of the bipartition of a network is that the matrix describing it is no longer a square matrix, but a rectangular one, because the number of nodes belonging to the two sets are often different. In this case the matrix is called *incidence matrix*. Let B be the incidence matrix describing a bipartite network composed by *i* groups and *j* original nodes, its elements B_{ij} are equal to:

$$B = \begin{cases} 1 & if\ node\ j\ belong\ to\ group\ i \\ 0 & otherwise \end{cases}$$

For instance, the 4×7 incidence matrix of the network shown in *Figure 2.4* is

$$B = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

From the incidence matrix we can also derive the matrices describing the two one mode projection of the bipartite network. Let i and k be two original nodes of a bipartite network described by the incidence matrix B; the product $B_{ij}B_{jk}$ is equal to 1 if and only if both i and k belong to the same group j. Then, the square matrix $P = B^T B$ is likely to an adjacency matrix of the one mode projection of original nodes. In the same way we define $P = B\,B^T$ as the matrix describing the one mode projection of groups.

Same properties defined about networks previously, like link direction, link weights or centrality, can be assigned to bipartite networks. About *degree centrality* we have to specify that in one mode networks we can compute the degree of a node summing over the row or over the column equivalently. In bipartite networks instead, summing over the rows we get the degree centrality score of original nodes and summing over the columns we get degree of groups. Let B be an incidence matrix composed of i-rows and j-columns, the degree centrality scores of groups and original nodes are:

$$degree\ centrality\ score\ of\ groups: \qquad k_i = \sum_{j=1}^{n} B_{ij}$$

$$degree\ centrality\ score\ of\ original\ nodes: \qquad k_j = \sum_{i=1}^{n} B_{ij}$$

For instance, the degree centrality scores of nodes belonging to the bipartite network illustrated in *Figure 2.4* are shown in *Table 2.3*.

node	degree centrality score
A	3
B	2
C	2
D	3
1	1
2	2
3	1
4	1
5	3
6	1
7	1

Table 2.3 *: Degree centrality scores concerning network in Figure 2.4.*

In contrast, *eigenvector centrality* definition explained previously cannot be applied to bipartite networks because the matrix must be square in order to compute its eigenvalues and eigenvectors. In order to bypass this limitation, we need to recall an important algebraic concept: the *Singular Value Decomposition* (SVD). According to the first SVD corollary (Golub & Van Loan, 2013) :

$$A^T A v_i = \sigma_i^2 v_i$$

$$A A^T u_i = \sigma_i^2 u_i$$

Where:

- A is a m-by-n matrix;
- σ_i are the *singular values* of A;
- u_i are the *left singular vectors* of A;
- v_i are the *right singular vectors* of A.

A strong connection exists between the SVD of a matrix A and the eigensystems of the square matrices $A^T A$ and AA^T (Golub & Van Loan, 2013). If A is the incidecidece matrix describing a bipartite network, the matrices $A^T A$ and AA^T represent the adjacency matrices of the two one-mode projection and the right and left eigenvectors can be used to rank original nodes and groups respectively.

2.2 Economic Complexity metrics

The aim of the Economic Complexity metrics is to find a way to characterize the socio-economic status of a country based on their export baskets composition (Hausmann, Hwang, & Rodrik, 2007), making an indirect measure of their intangibles capabilities (Pietronero, Cristelli, & Tacchella, 2013). We propose the following methods as an alternative to traditional economic competitiveness indices, capable of summarize the productive knowledge of countries from easy to get data. The Economic Complexity metrics have roots in the measures of centrality adopted in the Networks theory and they are able to rank both countries and products. The measures of Economic Complexity currently used are the *Method of Reflections* (MR) and the *Fitness and Complexity* (FC) algorithms. Despite of their common objective, these two methods differ not only in their approach to the problem but also in the obtained results. In this thesis work we present a new metric: the *GENeralized Economic comPlexitY index* (GENEPY), developed by a team from Politecnico di Torino (Sciarra, Chiarotti, Ridolfi, & Laio, 2020), that tries to reconcile the two previously mentioned metrics and furnishing a unique ranking for countries and products.

2.2.1 The country-product matrix

The network of Economic Complexity is a bipartite network, in which products *p* are connected to countries *c* only if these export them. Data concerning exports are often

quantified in USD (United States Dollars), thus the c×p incidence matrix describing the country-product network is weighted and its links represents the quantity in USD of product p exported by country c. Then, this matrix is modified according to the Balassa's definition of Revealed Comparative Advantage (RCA) (Balassa, 1965): the ratio between the quantity of product exported by a country and the quantity of the same product exported to the world market (Hausmann & Hidalgo, Country diversification, product ubiquity, and economic divergence, 2010). Analytically, we define RCA as:

$$RCA_{cp} = \frac{\frac{X_{cp}}{\sum_p X_{cp}}}{\frac{\sum_c X_{cp}}{\sum_{c,p} X_{cp}}},$$

where X_{cp} is the exports in dollars of the country c of a product p. Next, the incidence matrix have been modified furtherly: let M_{cp} be a generic element of this matrix identified by a given RCA value; we set M_{cp} equal to 1 if RCA is greater than 1, otherwise is equal to 0 (Hidalgo & Hausman, 2009). Through this process, we get a binary matrix representing a bipartite network in which the products are linked to the countries that export them only if they are relevant exporters (RCA>1).

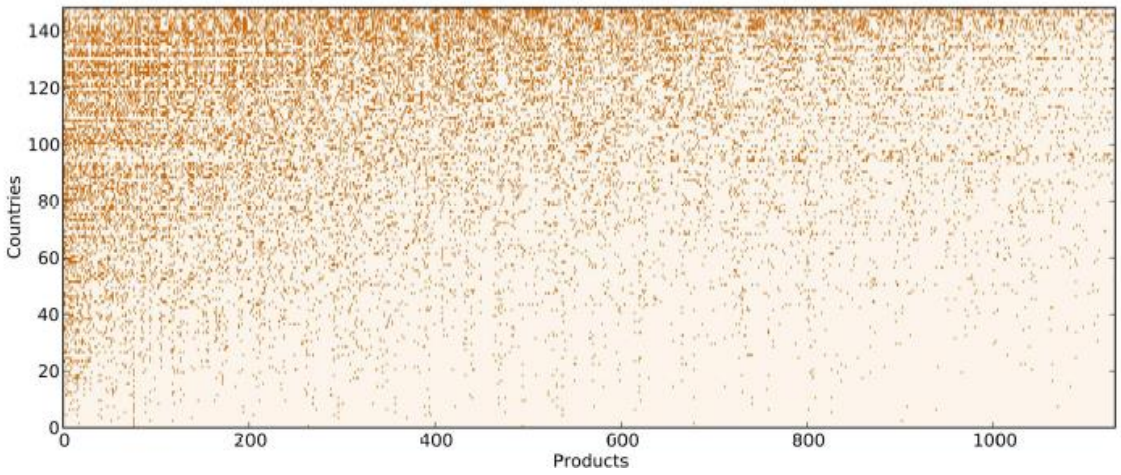


***Figure 2.5**: Graphical representation of Mcp for the year 2010. Reordering rows and columns by decreasing Diversity and Ubiquity it is evident the major triangular shape of the matrix* (Cristelli, Gabrielli, Tacchella, Caldarelli, & Pietronero, 2013).

2.2.2 Method of Reflections

Hidalgo and Hausmann, in their paper (Hidalgo & Hausman, 2009) introduced the first characterization of the previously described bipartite network. Because of the symmetric

structure of the bipartite network, they called their approach the *Method of Reflection* (MR). As we said before, the incidence matrix M_{cp} represents the economic network and its elements M_{cp} are equal to 1 if c is a significant exporter of product p, otherwise are 0. First, it is necessary to introduce two central concepts: the countries' *diversity* and the products' *ubiquity*. These are defined as the sum over rows and columns of the matrix M_{cp}, respectively.

$$Diversity = k_{c,0} = \sum_{p} M_{cp}$$

$$Ubiquity = k_{p,0} = \sum_{c} M_{cp}$$

The *Diversity* measures the number of different products a country is able to make and export; the *Ubiquity* measures instead the number of countries able to make a given product (Hausmann R. , Atlas of Economic Complexity, n.d.). We can think of Diversity and Ubiquity as the degree centrality scores of a bipartite network and we can use them as a first rough raking for countries and products. A graphical representation of Diversity and Ubiquity is given if *Figure 2.6.*

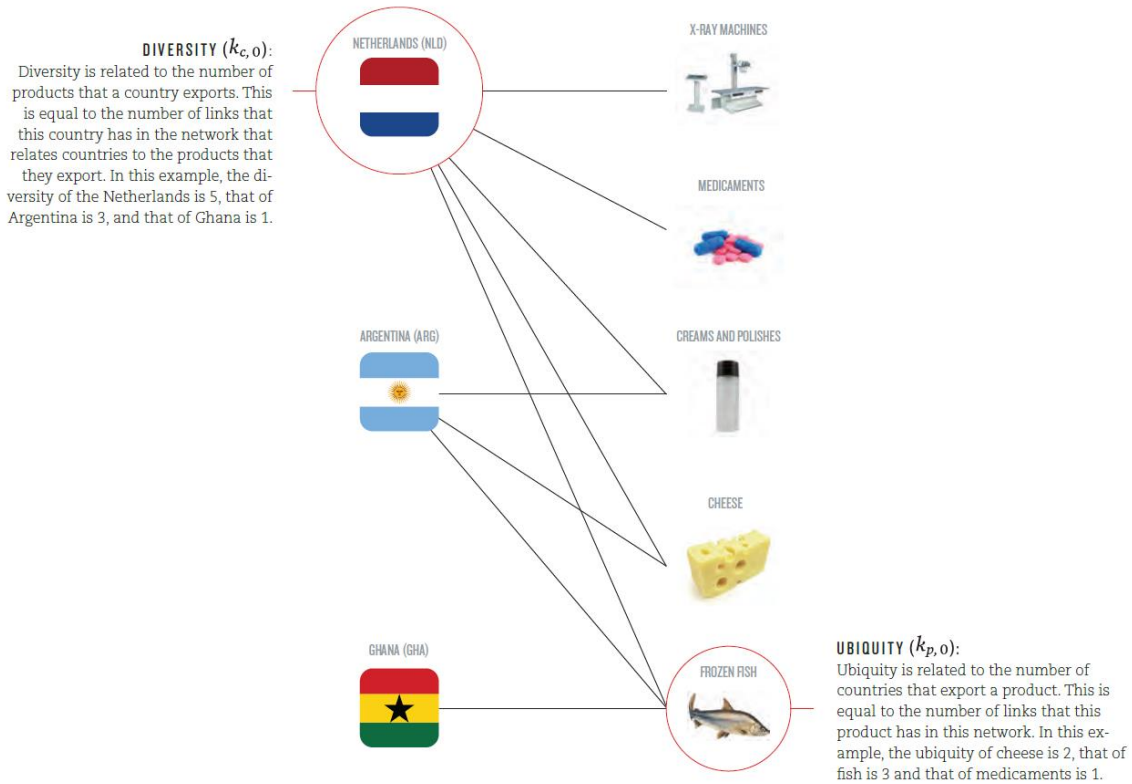


***Figure 2.6**: Graphical representation of diversity and ubiquity* (Hausmann R. , Atlas of Economic Complexity, 2013)

The algorithm developed by Hidalgo and Hausmann is built according to the assumption that Economic Complexity of Countries should be computed as the mean of the Complexity of products in their export basket, and in the same way the Complexity of a product correspond to the mean Complexity of the countries exporting them. We can summarize this concept by the recursion:

$$k_{c,N} = \frac{1}{k_{c,0}} \sum_{p} M_{cp} \cdot k_{p,N-1} \qquad (1)$$

$$k_{p,N} = \frac{1}{k_{p,0}} \sum_{c} M_{cp} \cdot k_{c,N-1} \qquad (2)$$

We then insert the second equation into the first one, to obtain:

$$k_{c,N} = \frac{1}{k_{c,0}} \sum_{p} M_{cp} \cdot \frac{1}{k_{c,0}} \sum_{c} M_{c'p} \cdot k_{c',N-2} \qquad (3)$$

$$k_{c,N} = \sum_{c'} M_{c'p} \cdot\cdot k_{c',N-2} \sum \frac{M_{cp} M_{c'p}}{k_{c,0} k_{p,0}} \qquad (4)$$

In addition, rewrite this as:

$$k_{c,N} = \sum_{c'} \widetilde{M_{cc'}} \, k_{c',N-2} \qquad (5)$$

Where

$$\widetilde{M_{cc'}} = \sum_{p} \frac{M_{cp} M_{c'p}}{k_{c,0} k_{p,0}} \qquad (6)$$

The iterative process stops when $k_{c,N} = k_{c,N-2} = 1$. This vector is a vector of 1s and precisely is the eigenvector associated with the largest eigenvalue of matrix $\widetilde{M_{cc'}}$. We cannot extract information from a vector of 1s, as consequence the eigenvector associated with the second largest eigenvalue carry most of the variance in the system and represents the measure of Economic Complexity (Simoes, Landry, & Hidalgo, n.d.).

Therefore, to rank countries, we define the *Economic Complexity Index* (ECI) as:

$$ECI = \frac{\vec{K} - average(\vec{K})}{stdev(\vec{K})} \qquad (7)$$

By inverting the process (substituting the (1) in the (2) instead of the (2) in the (1)), due to the symmetry of the problem, we can define the *Product Complexity Index* (PCI) as:

$$PCI = \frac{\vec{Q} - average(\vec{Q})}{stdev(\vec{Q})} \quad (8)$$

$\vec{Q}$ is the eigenvector of matrix $\widetilde{M_{pp'}}$ associated with the second largest eigenvalue.

Therefore, the approach of Hidalgo and Hausmann is reduced to a linear algebra eigenvalue problem with a classic iterative solution (Morrison, et al., 2017). In brief, the Method of Reflections measures the Economic Complexity of countries as the average of the Complexity of the exported products and vice versa (Figure 2.7). However, this way we lose information about countries' diversification and products' ubiquity (Sciarra, Chiarotti, Ridolfi, & Laio, 2020).

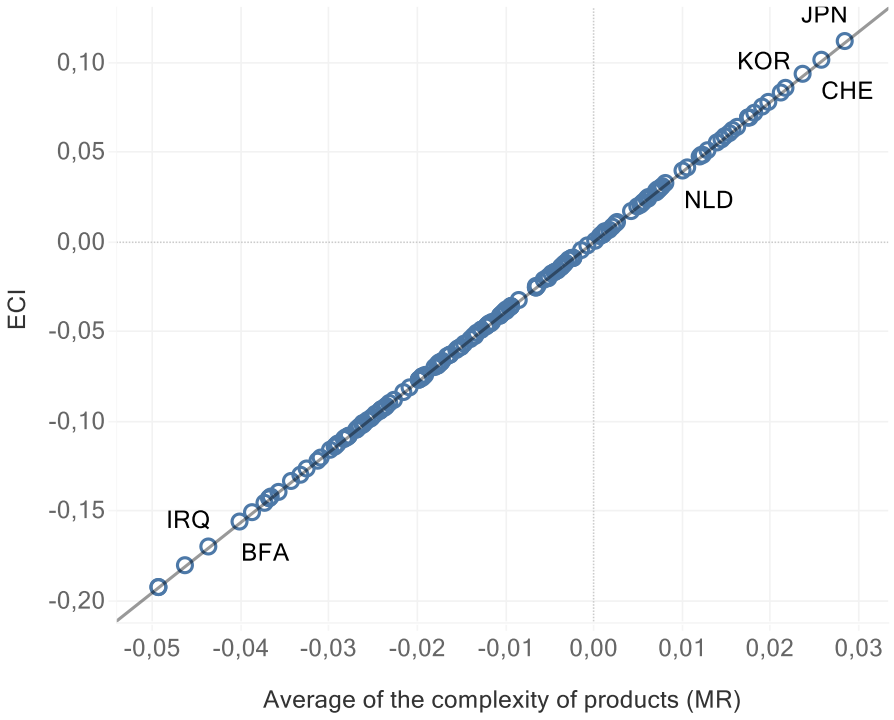


Figure 2.7: *Method of Reflections, year 2017. The Economic Complexity Index of countries corresponds the average of the complexity of the exported products.*

2.2.3 Fitness and Complexity algorithm

In 2012, Tacchella et al. proposed a new iterative non-linear approach for Economic Complexity evaluation, called Fitness-Complexity method (FC) (Tacchella, Cristelli, Caldarelli, Gabrielli, & Pietronero, 2012). According to them, the industrial development of a country, called *Fitness*, should be computed as the number of exported products weighted by their

Complexity (Pietronero, Cristelli, & Tacchella, 2013). Developed countries export a large ammount of products, so we cannot extract relevant information from the observation that a product is made by a developed country. Instead if we know that a less competitive country export a given product, as consequence the product's complexity should unavoidably downgrade. This crucial new concept involves non-linar distribution of the complexity of both countries an products.

Considering an unidirect bipartite network whose nodes are countries C and products P; as we have already seen, this type of network can be represented thtrough its incidence matrix, whose elements M_{cp} are equal to 1 if the quantity in USD$ of the product p exported from the country c is such as to have RCA>1, and zero otherwise. Tacchella et al. call *Fitness (F_c) of countr*ies and *Quality (Q_p) of products* the measures of Economic Complexity of countries and products respectively, defined by the following non-linear iterative map (Servedio, Buttà, Mazzilli, Tacchella, & Pietronero, 2018):

$$\begin{cases} F_c^{(n)} = \sum_{p'} M_{cp'} Q_{p'}^{(n-1)} & with\ 1 \leq c \leq C \\ Q_p^{(n)} = \left(\sum_{p'} M_{c'p} / F_{c'}^{(n-1)}\right)^{-1} & with\ 1 \leq p \leq P \end{cases} \qquad (9)$$

With initial values $F_c^{(0)} = Q_p^{(0)} = 1$ per $\forall\ c,p$. C and P are the total value of countries and exported products:

$$\sum_c F_c^{(n)} = C \qquad \sum_p Q_p^{(n)} = P \qquad (10)$$

The main properties of this new metric are: the Fittness of a country depends on the diversification weighted by exported products' Complexity (Pietronero, Cristelli, & Tacchella, 2013); the Complexity of a product inversely depends on the number of countries porducing it; if low developed countries can produce a certain product, its complexity is unavoidably degraded; the Fitness and Complexity distribution fit well the Pareto-like behaviour of monetary variables.

Sumarizing,the Fitness-Complexity algorithm is therefore also based on an adaptation of the eigenvector centrality, but with the addition of two postulate: the Quality of a product is lower if it is exported by more countries (Morrison, et al., 2017); the Fitness of a country is obtained as the sum of the qualities of the exported products (*Figure 2.8*), preserving the information related to the diversification of its export basket.

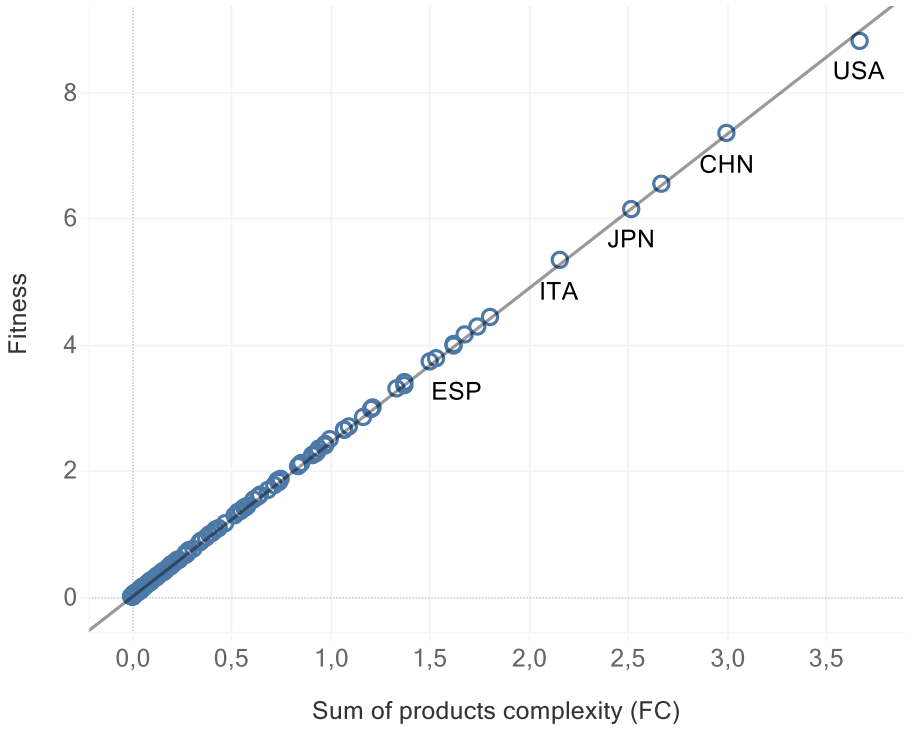


***Figure 2.8**: Fitness and Complexity, year 2017. The Fitness of countries corresponds the sum of the complexity of the exported products.*

2.2.4 Generalized Economic Complexity Index

The Genepy is a liqueur made from alpine herbs, typical of north-west Italy. Same way, Sciarra et al. propose to distil the information on economic complexity into a GENeralised Economic comPlexitY index (GENEPY) (Sciarra, Chiarotti, Ridolfi, & Laio, 2020). The GENEPY index, adopting a neat matematical outlook based on linear algebra application in bipartite networks, is able to reconcile the existing economic complexity metrics (Sciarra, Chiarotti, Ridolfi, & Laio, 2020).

This new metric is rooted in the different interpretation of centrality proposed by the autors themseves (Sciarra, Chiarotti, Laio, & Ridolfi, 2018): centrality can be seen as an intrinsic property of the nodes, thanks to which is possible to estimate the adjacency matrix of the network. Let G be an undirected, unweighted network described by the adjacency matrix A, whose elements Aij are euqal to 1 if i is linked to j, zero otherwise (Newman, Networks: An introduction, 2010). Let $\hat{A}$ be an estimator of the ajacency matrix; its values, $\hat{A}_{ij}$, are higher if nodes i and j share and edge and lower otherwise (Sciarra, Chiarotti, Laio, & Ridolfi, 2018). Accordind to Sciarra et al. view, the estiamtor of a generic element of the adjacency matrix A_{ij}

is conditioned by some arising properties x_i and x_j of the node *i* and *j* respectively, in formulas $\hat{A}_{ij} = f(x_i, x_j)$. The values x_i are established by minimising the sum of the squared (SE) differences between the element and its estimator:

$$SE = \sum_i \sum_j (A_{ij} - \hat{A}_{ij})^2 = \sum_i \sum_j (A_{ij} - f(x_i, x_j))^2 \qquad (11)$$

Considering the symmetry of the matrix, the minimization procedure, with respect to x_i, corresponds to the solution of the following equation

$$\frac{\partial SE}{\partial x_i} = 4 \sum_j [A_{ij} - f(x_i, x_j)] \cdot \frac{\partial f(x_i, x_j)}{\partial x_i} = 0 \qquad (12)$$

This equation concerns just node i, but considering a set of N nodes, N equations are obtained; allowing one to estimate the centrality score of all the N nodes (Sciarra, Chiarotti, Laio, & Ridolfi, 2018). Thus, a node *i* is more important than a node *j* if, when its property is excluded from the estimation process, the *SE* change more than if excluding the property of the node *j*. The quantitative measure of the effect a variable has, during an estimation procedure, is called *unique contribution* (Nathas, Oswald, & Nimon, 2012). The unique contribution (UC) of node i is defined as:

$$UC_i = R_N^2 - R_{N\setminus i}^2 = \frac{SE_{N\setminus i} - SE_N}{TSS} \qquad (13)$$

Where $R^2 = 1 - \frac{SE}{TSS}$ is the coefficient of determination (the subscripts N\i refers to the case when property x_i is not considered in the estimation procsess, indeed subscript N take into account all the x_i values during the estimation), in which TSS (Total Sum of Squares) is equal to $\sum_i \sum_j (A_{ij} - \bar{A})$. The importance (centrality) of a node is thus reflected by higher value of UC. Moreover, by defining the function *f* it is possbile to obtain different centrality metrics (like degree centrality or eigenvector centrality, as shown in *Table 2.3*) (Sciarra, Chiarotti, Laio, & Ridolfi, 2018).

Undirected networks			
Estimator function *f*	**Centrality of node *i***	**Unique contribution of node *i***	**Corresponding metric**
$f_1 = \frac{K_{tot}}{N}\left(x_i + x_j - \frac{1}{N}\right)$	$x_i = \frac{k_i}{K_{tot}}$	$UC_i = \frac{2(N+1)k_i^2}{N^2\, TSS}$	Degree centrality
$f_2 = \gamma x_i x_j$	$x_i = \frac{1}{\gamma}\sum_j A_{ij} x_j$	$UC_i = \frac{\gamma x_i^2}{TSS}(\gamma x_i^2 + 2\gamma)$	Eigenvector centrality

***Table 2.4**: Examples of the estimator function* (Sciarra, Chiarotti, Laio, & Ridolfi, 2018).

Extenting this new perspective to *multi-component* centrality measures means tanking into account more than one scalar property in describing network's nodes relevance. Analitically,

this equals to consider x_i as an s-dimensional vector, where s are the properties of the node influencing the estimation process. For instance, adopting the function f_2 (*Table 2.3*), the estimator takes the folowing form:

$$\hat{A}_{ij}(s) = \gamma_1 x_{i,1} x_{j,1} + \cdots + \gamma_k x_{i,k} x_{j,k} + \cdots + \gamma_s x_{i,s} x_{j,s} \qquad (14)$$

It can be shown (Sciarra, Chiarotti, Laio, & Ridolfi, 2018) that γ_k is the k-th eigenvalue of the adjacency matrix and x_k its corresponding eigenvector, recalling a solution that corresponds to the *Singular Value Decomposition* (SVD) (Golub & Van Loan, 2013) of the original matrix stopped at the order s, and providing the following expression of UC extended to the multi-dimensional case

$$UC_i(s) = \frac{1}{TSS}\left[\left(\sum_{k=1}^{s} \gamma_k x_{i,k}^2\right)^2 + 2\sum_{k=1}^{s} \gamma_k^2 x_{i,k}^2\right] \qquad (15)$$

By tanking into account more than one dimension, allows node centrality ranking to be more sophisticated, embedding node's features that one-dimension does not consider (for instance eigenvectors beyond the first embeds informations about structure and clustering of the network (Newmann, 2006)).

The transposition of what already shown in the context of economic complexity gave birth to GENEPY. Given a bipartite network composed of C countries and P products, the theories of economic complexity aim to determine two network properties, the complexity of countries (X_c) and the complexity of products (Y_p), by a system of coupled equations caracterized by two linear function, f and g:

$$\begin{cases} X_c = f(Y_1, Y_2, \dots, Y_p, M_{cp}) & p = [1, \dots, P] \\ Y_p = g(X_1, X_2, \dots, X_c, M_{cp}) & c = [1, \dots, C] \end{cases} \qquad (16)$$

The linearity of the functions allows to bring the solution of the system to the solution of an eigenvalue problem. Let W be an appropriate transformation matrix, whose elements, W_{cp}, are derived from the incidence matrix M. Defined λ as the eigenvalue of the mentioned eigen-problem, the system become:

$$\begin{cases} X_c = \frac{1}{\sqrt{\lambda}} \sum_p W_{cp} Y_p \\ Y_p = \frac{1}{\sqrt{\lambda}} \sum_c W_{cp} X_c \end{cases} \qquad (17)$$

An adequate definition of W_{cp} allows us to derive from the system the two previously described metrics of economic complexity (*Figure 2.9* and *Figure 2.10*), MR and FC: by setting $W_{cp} = M_{cp}/\sqrt{k_c k_p}$ the system turns into the MR; by setting $W_{cp} = M_{cp}/k_c k'_p$, $X_{c,1} = F_c/k_c$ and $Y_{p,1} = Q_p k'_p$ (where $k'_p = \sum_c M_{cp}/k_c$) the system turns into the FC method. The concept of multidimensional complexity enables to synthetize the two metrics into a single metric of centrality, called GENEPY and defined for countries as follows:

$$GENEPY_c = \left(\sum_{i=1}^{2} \lambda_i X_{c,i}^2\right)^2 + 2\sum_{i=1}^{2} \lambda_i^2 X_{c,i}^2 \qquad (18)$$

$$\begin{cases} N_{cc*} = \sum_P W_{cp} W_{c*p} = \sum_p \dfrac{M_{cp} M_{c*p}}{k_c k_{c*} \left(k'_p\right)^2}, \ \ if\ c \neq c^*, \\ N_{cc*} = 0, \ \ if\ c = c^* \end{cases} \qquad (19)$$

where λ_1 and λ_2 are the first two largest eigenvalues and $X_{c,1}$ and $X_{c,2}$ are the eigenvector associated to the two largest eigenvalues of the proximity matrix N. The proximity matrix is a squared symmetric matrix defined as $N=WW^T$ for the countries (and $G=W^TW$ for products). We can better understand the way GENEPY distils the information by looking at its components, the two eigenvectors $X_{c,1}$ and $X_{c,2}$: the elements of the first eigenvector represent the eigenvector centrality of the countries; the elements of the second eigenvector carry information about the clustering of countries that share capabilities among them. So, GENEPY, by combining the advantages of the two existing metrics of economic complexity, identifies that set of capabilities a country possesses and shares with others (Sciarra, Chiarotti, Ridolfi, & Laio, 2020). About the ranking of products, it is enough to consider matrix $G=W^TW$ instead of matrix $N=W^TW$.

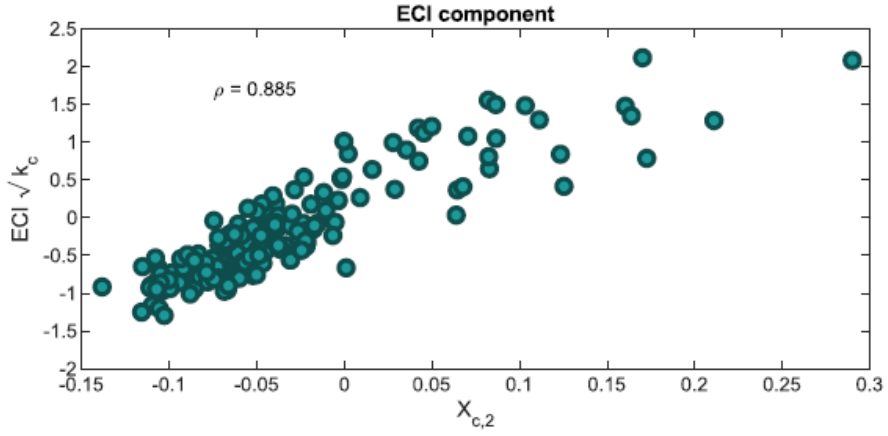


Figure 2.9*: Scatter plot of the second component of GENEPY, $X_{c,2}$, correlated to the ECI values over the countries squared degree - year 2017 (Sciarra, Chiarotti, Ridolfi, & Laio, 2020).*

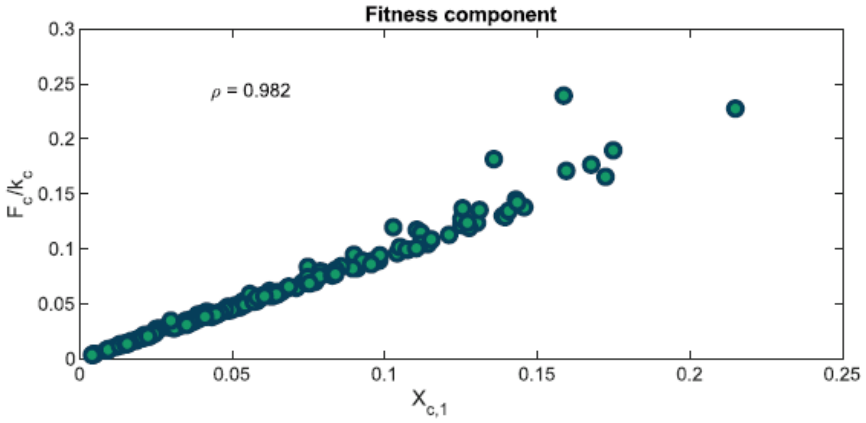


Figure 2.10: *Scatter plot of the first component of GENEPY, $X_{c,1}$, correlated to the values of Fitness over the countries degree (year 2017)* (Sciarra, Chiarotti, Ridolfi, & Laio, 2020).

CHAPTER 3

3 Materials and Methods

The EC algorithms described in the previous chapter need specific inputs in order to be performed, specifically the inputs are the incidence matrices describing the structure of the bipartite economic network. The trade data, containing annual records of exported products, are required to perform the network's structure characterization procedure, accumulating quantities of exported commodities during each year. Data has been processed creating appropriate MATLAB scripts, because trade data needed to be cleaned and organized as incidence matrix. Moreover, in order to investigate causation between results obtained by EC metrics and GDP per capita, in this chapter we describe a technique, called Convergent Cross Mapping, thanks to which we have been able to perform this task. Further details follow.

3.1 Materials

Each year every country reports their trade flows to the Statistical division of the United Nations, which collects and disseminates them through the COMTRADE: the Commodities Trade Statistics database. Both exporters and importers report their trade flows to United Nations, so the data are double recorded. For instance, trade from country A to country B may be reported both by A (as an export to B) and B (as an import from A). Theoretically this two values should be the same, but in practice, for two reasons, are never identical. The first reason is that import values are reported CIF (cost, insurance and freight), whereas exports are free on board; the second reason is the presence of errors due to the uncertain destination on some exports or the incorrect classification of product.

The export data used in this thesis are taken from the BACI-CEPII dataset (http://www.cepi.fr/). The BACI World Trade Database, supporting all the analyses

conducted in this work, is the result of a reconciliation procedure aimed at achieving a consistent single flow (Gaulier & Zignago, 2009). This dataset covers more than 200 countries and 5000 products classified through the Harmonized Commodity Description and Coding Systems (HS). The HS is one of the most used nomenclature for classification of trade goods according to a six-digit code (UN International Trade Statistics, 2017), representing the most detailed classification at the international level (Gaulier & Zignago, 2009). Each product is identified by a six digits code made up of three parts representing the structure of the HS itself:

- the first two digits represent the chapter (from 0 to 99);
- the next two digits identify headings;
- the last two digits identify sub-headings.

As the number of digits increases, the detail level of the product description increases. For instance, HS code 100630 consist of Chapter 10 (Cereals), Heading 06 (Rice), and Subheading 30 (Semi-miller or wholly milled rice, whether or not polished).

The first version of the HS dates back to 1988, but over the years it has been periodically updated. In this work, we are referring to the 1992 version (HS92) and to the 4-digit classification: obtained by removing the last two digits and grouping products only according to the first four, this reduce the number of products to approximately 1200. Macroscopically, we organize products in sixteen categories linked to the first two digits (*Table 3.1*).

First two digits	Category
01 to 05	Animals
06 to 15	Vegetables
16 to 24	Food Products
25 to 26	Minerals
27	Fuels
28 to 38	Chemicals
39 to 40	Plastic or Rubber
41 to 43	Hides and Skins
44 to 49	Wood
50 to 63	Textiles and Clothing
64 to 67	Footwear
68 to 71	Stone and Glass
72 to 83	Metals
84 to 85	Machinery and Electronics
86 to 89	Transportation
90 to 99	Miscellaneous

Table 3.1: *Products' Categories.*

Data concerning exports, even if reviewed in the BACI dataset, needs a cleaning procedure consisting of:

- unification of export data concerning repeated country-product pairs;
- elimination of products which are not exported from any country;
- elimination of countries which do not export any product;
- elimination of countries whose total exports are less than $10^{-5}$of global exports;
- removal of Taiwan (also known as Republic of China) because it has not been officially recognized as country.

After this cleaning process we have been ready to shape trade data into the incidence matrix and run the EC algorithms.

3.1.1 Input matrices

In paragraph 2.2.1 we said that: given C countries and P products, the trade data concerning export ore organized in a $C \times P$ incidence matrix whose elements are equal to 1 if country c is a relevant exporter (if the Balassa's definition of RCA is higher than 1) of product p, otherwise are equal to zero. Matrices of this type are called binary matrices and are used to describe unweighted networks. Anyway, the EC algorithms are able to rank countries and products even if the matrix is not binary and thus the network is weighted. One of the aim of this thesis is to investigate the relationship between the results obtained adopting different matrices as input for the EC metrics.

The used input matrices are :

- The USD matrix, whose elements are equal to the quantity, expressed in United States Dollars, of a product exported by a given country (the network is weighted);
- The RCA matrix, whose elements are equal to the RCA value, according to the Balassa's definition of RCA (the network is weighted);
- The binary matrix, whose elements are equal to 1 or 0 if the relative RCA is respectively higher or lower than 1 (the network in unweighted).

3.2 Detecting causation in Complex Systems

As first suggested by Sugihara et al. (Sugihara, et al., 2012), in order to investigate causation between the Economic Complexity metrics and the Gross Domestic Product per capita we used a technique called Convergent Cross Mapping (CCM), able to compute causality correlation between time series belonging to the same dynamic system. Thus, results obtained from EC algorithms over years became, together with GDP time series, input for the CCM algorithm.

Complex systems often have nonlinear dynamic behaviour. Nonlinear systems are systems that is possible to model by nonlinear algebraic and/or nonlinear differential equations (Nayfeh & Balachandran, 2008), opening possibilities for complex behaviour that are not possible in linear systems. In these systems there is no proportionality and no simple causation between the response and their input, meaning that small changes can have surprising and unexpected effects (Willy, Neugebauer, & Gerngrob, 2003). One of the classics examples of modern nonlinear dynamics is the *Lorentz attractor*, represented by a system of differential equations capable of generating chaotic behaviour. An attractor is a set towards which a dynamic system evolves after a sufficiently long time.

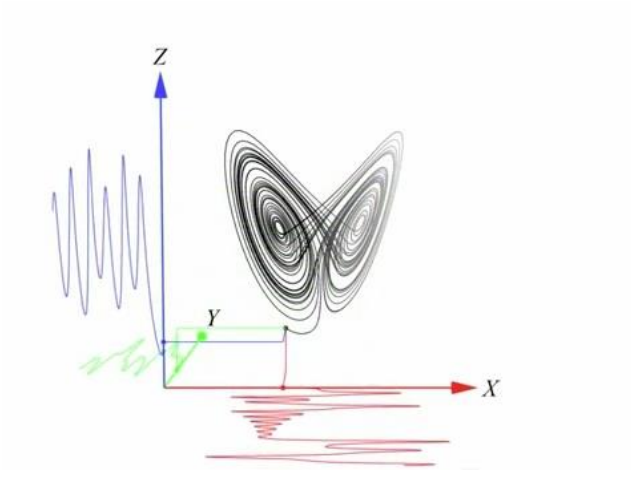


***Figure 3.1**: Lorentz attractor and its coordinate time series projections* (Sugihara, et al., 2012).

In systems in which dynamical variables cannot be monitored simultaneously or are infinite in number, the search of this attractor is problematic (Huke, 2006). In 1981, Takens (Takens, 1981) showed how it is possible to reconstruct the attractor in the *phase space* (in dynamical system theory, a phase space is a space where all possible states of a system are embedded) of a system through the time series of a single observable variable. The technique described

by Takens is called *method of delays*, because this technique allows to reconstruct a shadow version of the original system just by using lags of a single time series. The method proposed by Takens shows a *one-to-one mapping* between the attractor of the original system and its reconstruction, preserving its fundamental mathematical features.

According to Takens it is feasible to reconstruct a shadow manifold of a system's phase space through lags of a single observable variable.

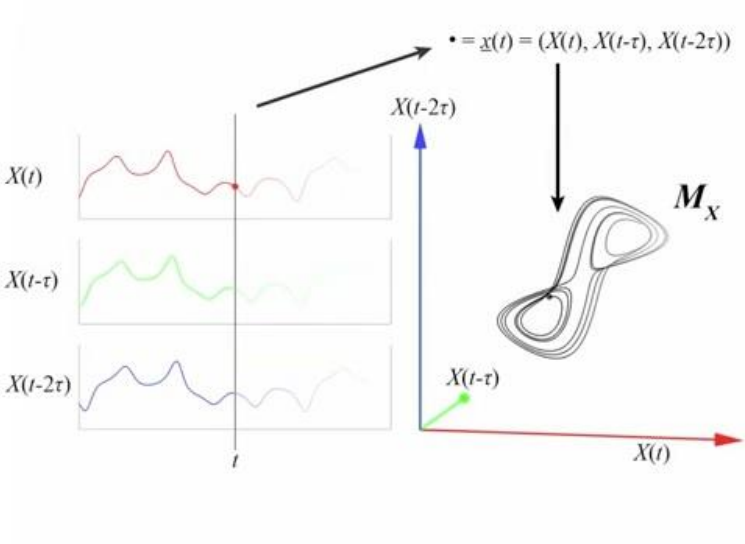


***Figure 3.2**: Shadow manifold of Lorentz attractor* (Sugihara, et al., 2012)

The same observation can be made considering another variable belonging to the same dynamic system, obtaining a reconstruction similar to the one obtained by the first variable (Sugihara, et al., 2012). For instance, let us consider the canonical Lorentz's system: let M be the phase space of the system; let X and Y be two variables belonging to this system and let M_x and M_y be the phase space reconstructed from variable X and Y respectively. Since both M_x and M_y are able to map one-to-one M, they also can be mapped one-to-one mutually. This allows us to use the time series of Y to estimate X and vice-versa, a technique called *Cross Mapping*. As the size of the time series increases, the accuracy of the estimation increases: this phenomenon is called *Convergent Cross Mapping* (CCM) and it is used as a practical method to identify causation (Sugihara, et al., 2012).

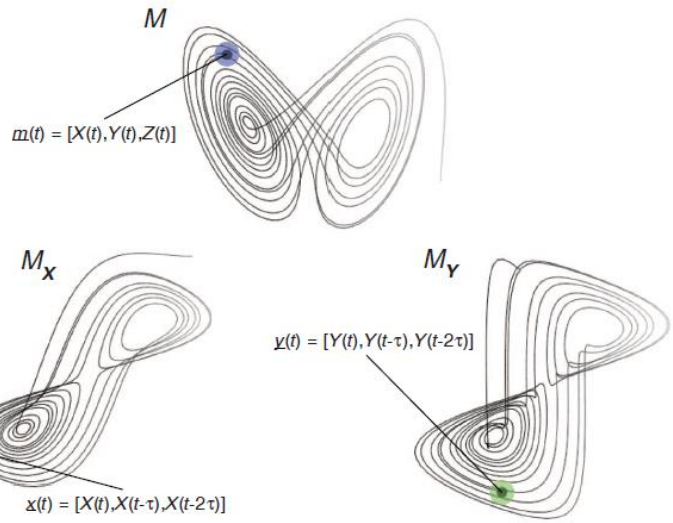


Figure 3.3 *: Convergent Cross Mapping based on the canonical Lorenz system in X,Y and Z* **(Sugihara, et al., 2012)**.

3.2.1 Convergent Cross Mapping algorithm

Let X(t) and Y(t) be two time series of length L. The algorithm computing the Convergent Cross Mapping, between the time series, may be divided into 5 steps (McCracken & Weige, 2014):

1. Creating the set of the lagged-coordinate vectors of X, called *shadow manifold* (M_x);
2. Search of the nearest neighbours, at time t, to a point in the shadow manifold;
3. Creating weights according to the distance from nearest neighbours;
4. Estimation of Y using weights (called $\hat{Y}(t)|M_x$);
5. Computing the correlation between original time series (Y) and the estimated one ($Y|M_x$).

Further details follow.

3.2.1.1 Shadow manifold creation

Let *{X}={ X(1), X(2), … , X(L) }* and *{Y}={ Y(1), Y(2), … , Y(L) }* be two time series of length L. We begin by building the shadow manifold of X, which is the set of lagged-coordinate vectors (Sugihara, et al., 2012):

$$\underline{x(t)} = < X(t), X(t-\tau), X(t-2\tau), \dots\dots\dots X(t-(E-1)\tau) >$$

Where E is the embedded dimension (in other words, E is the dimension of the reconstruction) and τ is the time step delay for the reconstruction. The first vector is created at $t = 1 + (E - 1)\tau$ and the last vector is created at $t = L$. This set of vectors is the shadow manifold M_x (Sugihara, et al., 2012).

3.2.1.2 Finding the nearest neighbours

We need to consider a minimum number of E+1 nearest neighbours in an E dimensional space (Sugihara, Grenfell, & Mccreddie May, Distinguishing error from chaos in ecological time series, 1990). In order to generate an estimation of Y(t) , we begin by finding the E+1 nearest neighbour of the lagged-coordinate vector, $\underline{x(t)}$, on M_x. For each $\underline{x(t)}$ the nearest neighbour search provides also a set of distances ordered from closest to farthest $\{d_1, d_2, \dots, d_{E+1}\}$ and computed as Euclidean distances between vectors.

3.2.1.3 Creating weights

This step consists in computing for each nearest neighbour an associated weight (w_i) depending on distances. In formulas:

$$w_i = \frac{u_i}{\sum_{j=1}^{E+1} u_j}$$

In which, $u_i = e^{-d_i/d_1}$ and i is the i-th nearest neighbour on M_x.

3.2.1.4 Time series estimation

Adopting the weights calculated above, we are able to estimate a point Y(t) in {Y} using a locally weighted mean of the E+1 values:

$$\hat{Y}(t)|M_x = \sum_{i=1}^{E+1} w_i Y(t_i)$$

Where i is the i-th nearest neighbour on M_x and $Y(t_i)$ are the contemporaneous values of Y (Sugihara, et al., 2012).

3.2.1.5 Computing the correlation

The CCM correlation is computed as:

$$C_{YX} = [\rho(Y, Y|M_x)]$$

In which ρ is the standard Pearson's correlation coefficient between the original time series and the one estimated through the cross-mapping technique. The C_{YX} value tell us information about how well X may be used to estimate Y. Obviously, it is also possible use Y to estimate X just by inverting X and Y in the previously described steps.

3.2.1.6 CCM code

In practice, to perform our CCM analysis we used the Matlab code realized by Jozef Jakubik (Krakovskà, Jakubìk, Budacova, & Holecyova, 2016) and based on Sugihara's algorithm. Additionally, we developed Matlab scripts able to repeat the following function at increasing time series length or time delays (details in the next chapter).

```
1. function [ SugiCorr , SugiY , SugiX , origY , origX ] = Sugi( X ,
   Y , tau , E )
2.
3. % Calculating Sugihara's CMM.
4. %
5. % References:
6. % Sugihara, George, et al., Detecting Causality in Complex
   Ecosystems, Science 26 October 2012, Vol. 338, no. 6106, pp. 496-
   500.
7. %
8. % Inputs:
9. % X,Y - time series with the same length
10.  % tau - time step for the reconstruction
11.  % E   - dimension of the reconstruction
12.  % The number of neighbourhoods for Sugihara's CCM method is E+1
13.  %
14.  % Outputs:
15.  % SugiCorr - correlation between the CCM estimation of original
   data and original data
16.  % SugiY, SugiX - the CCM estimate of original data
17.  % origY, origX - original data
18.
19.  L=length(X);
20.  T=1+(E-1)*tau;
21.  Xm=zeros((L-T+1),E);
22.  Ym=zeros((L-T+1),E);
23.  SugiN=E+1;
24.  N = L-T+1;
25.
26.  %% RECONTRUCTIONS OF ORIGINAL SYSTEMS
27.
28.  for t=1:(L-T+1)
29.      Xm(t,:)=X((T+t-1):-tau:(T+t-1-(E-1)*tau));
```

```
30.       Ym(t,:)=Y((T+t-1):-tau:(T+t-1-(E-1)*tau));
31.   end
32.   %%
33.
34.   SugiX=zeros(N,1);
35.   SugiY=zeros(N,1);
36.
37.   origY=Y(T:end);
38.   origX=X(T:end);
39.
40.   parfor j=1:N
41.
42.   %% neighbourhood search
43.
44.   [n1,d1]=knnsearch(Xm,Xm(j,:),'k',E+2);
45.   [n2,d2]=knnsearch(Ym,Ym(j,:),'k',E+2);
46.   susY=origY(n1(2:end));
47.   susX=origX(n2(2:end));
48.
49.   %% CMM
50.
51.   SugsusY=susY(1:SugiN);
52.   SugsusX=susX(1:SugiN);
53.   Sugid1=d1(:,2:SugiN+1);
54.   Sugid2=d2(:,2:SugiN+1);
55.   u1=exp(-Sugid1./(Sugid1(:,1)*ones(1,SugiN)));
56.   u2=exp(-Sugid2./(Sugid2(:,1)*ones(1,SugiN)));
57.   w1=u1./(sum(u1,2)*ones(1,SugiN));
58.   w2=u2./(sum(u2,2)*ones(1,SugiN));
59.   SugiY(j)= w1*SugsusY;
60.   SugiX(j)= w2*SugsusX;
61.
62.   end
63.
64.   SugiCorr1=corrcoef(origY,SugiY);
65.   SugiCorr(2,1)=SugiCorr1(1,2);
66.
67.   SugiCorr2=corrcoef(origX,SugiX);
68.   SugiCorr(1,1)=SugiCorr2(1,2);
69.
70.   end
```

CHAPTER 4

4 Results

The 4-digit level matrices regarding trade data since 1995 to 2017 have been used as input for the before mentioned Economic Complexity metrics. We used MATLAB, an environment for numerical calculation and statistical analysis, in order to shape trade data into incidence matrices, and then run the algorithms. TABLEAU has also been a powerful software thanks to which we have been able to make interesting visualizations regarding complexity of countries and products.

Our first goal is to analyse the three mentioned EC metrics, visualising the main differences and similarities among them and focusing on the evolution over time of the complexity of both countries and products. Particular attention has been paid to the complexity of products, because most of the papers on Economic Complexity focused on the reliability of the results looking mainly at countries' competitiveness.

The second goal is to investigate the causal relationship between the EC metrics and the most currently used economic competitiveness index, the Gross Domestic Product per capita (GDPpc). In order to do this, we used the time series of Complexity and GDP per capita for each country as input to compute their causality correlation through the CCM method.

4.1 Complexity of countries

The ability of Economic Complexity metrics to capture the level of industrial sophistication of countries can be assessed initially by observing the temporal evolution of the ranking of countries' complexity. The following figures illustrate how the ranking changed from 1995 to 2015, adopting a time interval of 5 years. Looking from *Figure 4.1* to *Figure 4.3,* we have a preliminary comparison between the results obtained with the 3 different algorithms. The

Method of Reflections (*Figure 4.1*) is characterized by a fairly stable trend about the highest positions, Japan (JPN) is the undisputed leader during whole time. Japan is a high-income country, its industrial diversification covers practically all sector of production, but according to the data retrieved from the *World Bank* (data.worldbank.org), its Gross Domestic Product per capita is not high enough to rank it among the top 20 countries in 2015. From the point of view of Hidalgo and Hausmann (Hausmann R. , Atlas of Economic Complexity, n.d.) the fact that a country is more complex than its actual income level means that it is expected to grow in the future, but we cannot yet consider this statement as an absolute truth.

As shown in *Figure 4.2* and *4.3,* the Fitness and Complexity algorithm and GENEPY agree more or less faithfully with the ranking obtained by the Method of Reflections about the top positions. Even though FC ranking differ more from MR than GENEPY; in fact, Germany (DEU) is the top ranked country, over most of the time, according to Fitness scores.

The behaviour of countries that in recent years have considerably increased their economic power is much more interesting. Countries like China (CHN), Korea (KOR), Hong Kong (HKG) or Singapore (SGP) would be a good example. GENEPY is the algorithm that best evaluates the industrial competitiveness of these countries, ranking them within the first six positions in 2015 (*Figure 4.3*). China is heavily penalized by the MR: the algorithm developed by Hidalgo e Hausmann takes into account the well-known economic growth of this country showing increasing complexity values (*Figure 4.4*), but the ECI of China is never high enough to rank it in in the top 20 positions.

It is interesting to note that the FC method, which has been praised for how enhances China's economic growth, does not rank a country like Singapore, which has been characterized by a notable economic growth in recent years, in the top 20 positions despite the fact that in 2015 it was ranked as the ninth country for Gross Domestic Product per capita by the World Bank.

On the basis of these preliminary views we can therefore argue that GENEPY is the metric that best considers the economic competitiveness of countries that certainly are in a growing phase. However, more in depth analysis are required in order to state this.

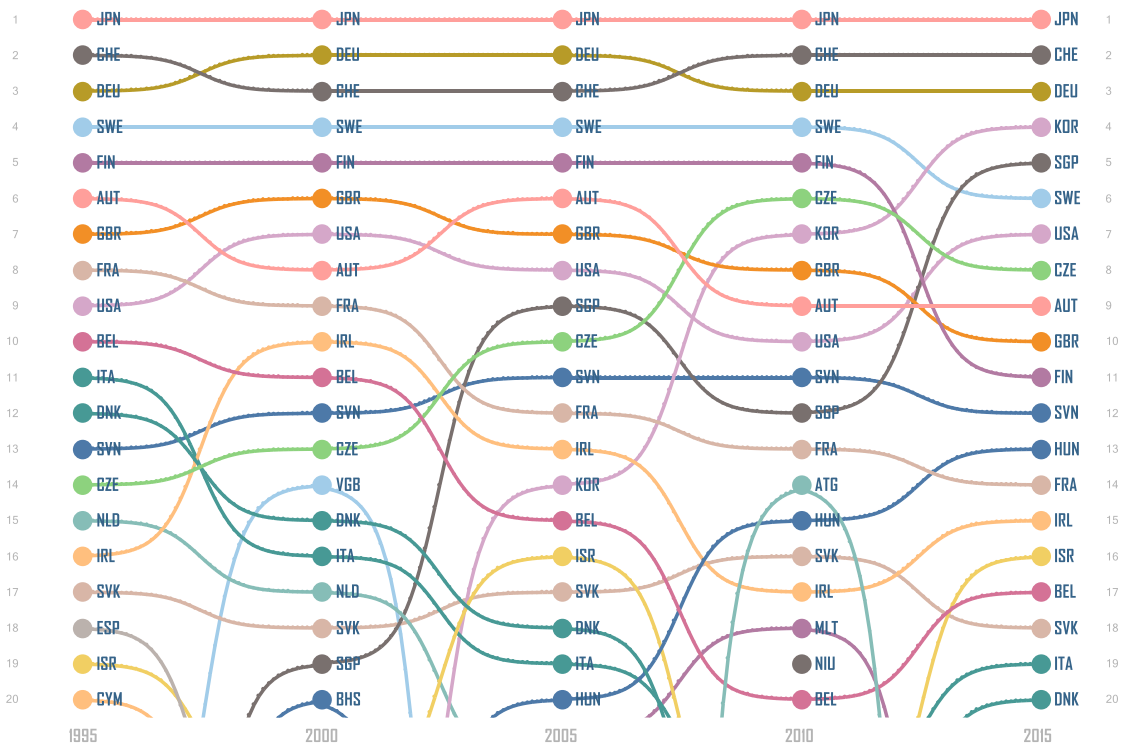


***Figure 4.1**: Ranking of countries, ordered by their ECI values, from 1995 to 2015.*

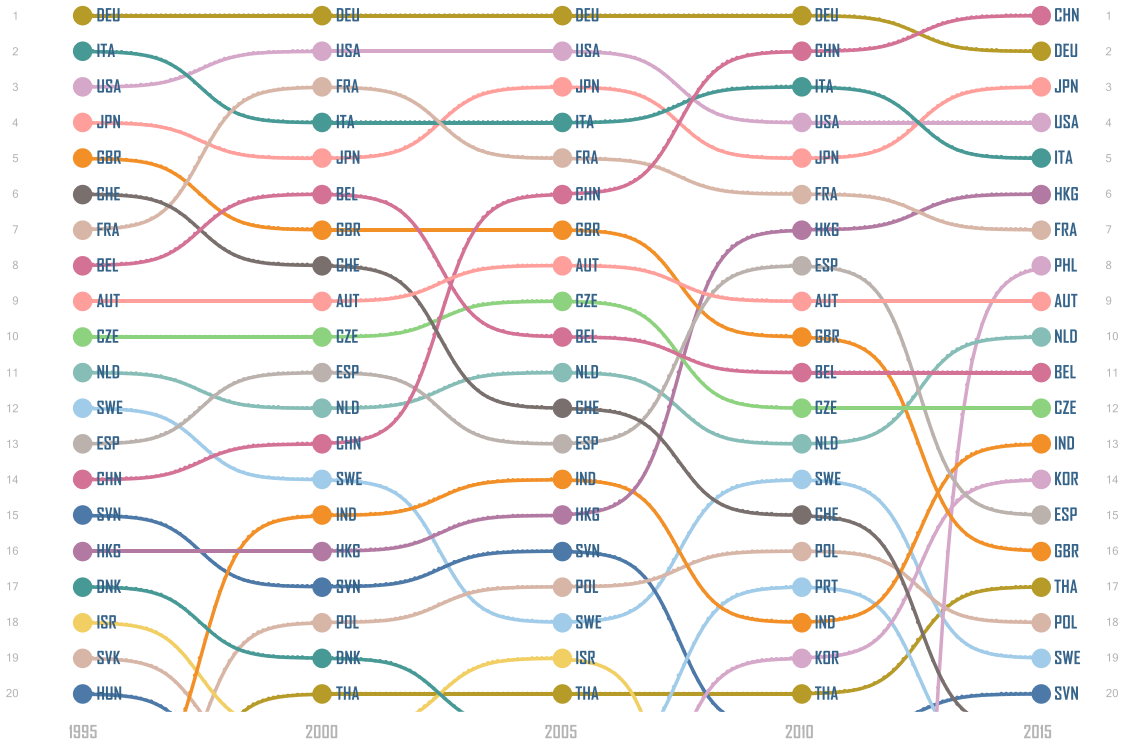


***Figure 4.2:** Ranking of countries, ordered by their Fitness values, from 1995 to 2015.*

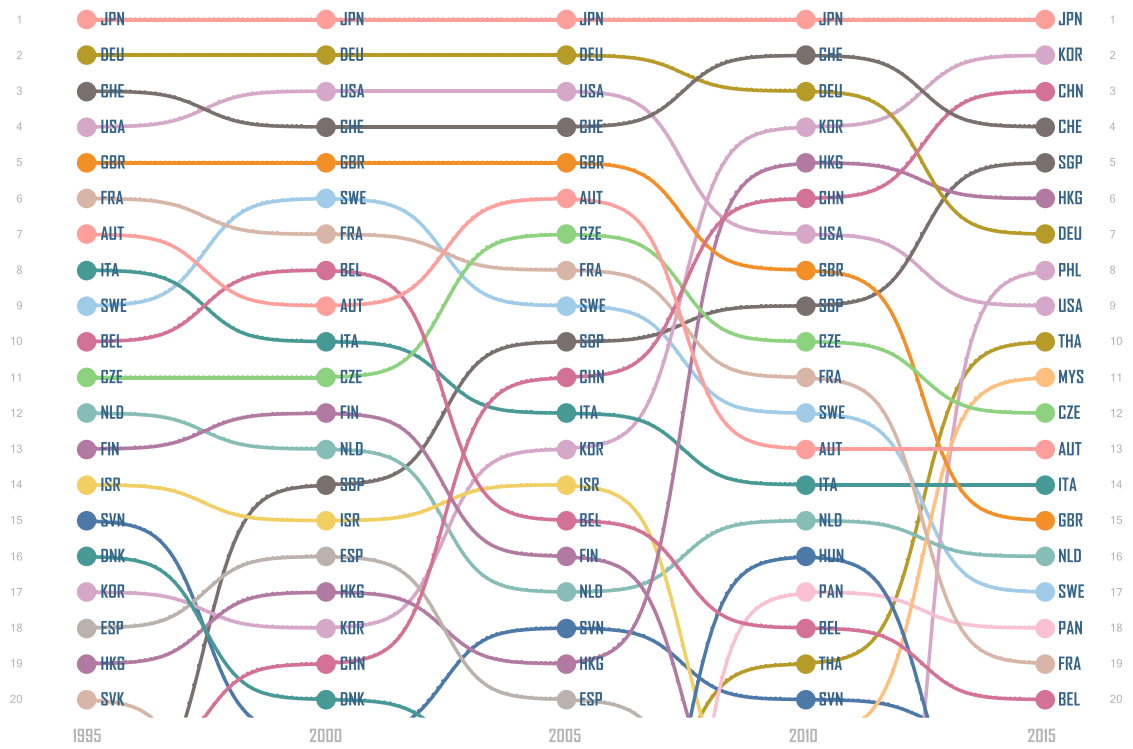


Figure 4.3: *Ranking of countries, ordered by their GENEPY values, from 1995 to 2015.*

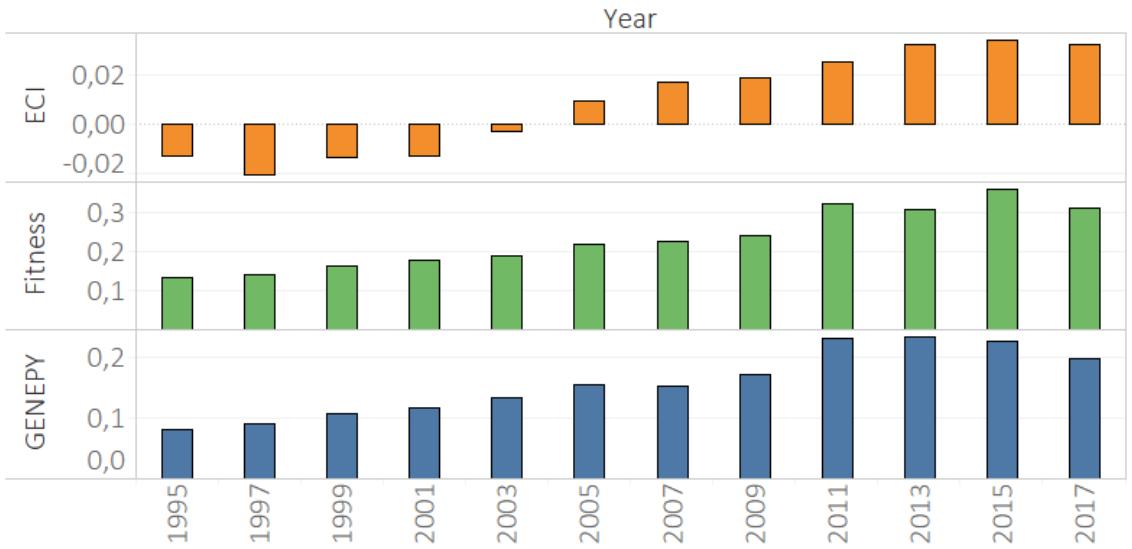


Figure 4.4*: Evolution of China's complexity from 1995 to 2015. Note that the values are normalized through the Frobenius' norm, in order to avoid excessive disparity between the results obtained with the different methods.*

In order to have a quantitative measure of how much each EC method differs from the others in classifying countries, we calculated the *Pearson's correlation coefficient* ($\rho_{X,Y}$) by matching the complexities' results for 2017 (*Table 3.1*). The Pearson's correlation coefficient is defined as follows:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Where $cov(X,Y)$ is the covariance, σ_X is the standard deviation of X and σ_Y is the standard deviation of Y.

Y	X	Pearson's correlation coefficient
Fitness	ECI	0.73
GENEPY	ECI	0.79
GENEPY	Fitness	0.89

***Table 4.1**: Pearson correlation coefficient between Complexity of countries calculated by the three different EC metrics for the year 2017.*

For the same year we also show the scatter plots (*Figure 4.5* to *4.7*), regarding results obtained by alternatively coupling the different EC metrics, in order to have a graphic feedback of what is shown in *Table 4.1*.

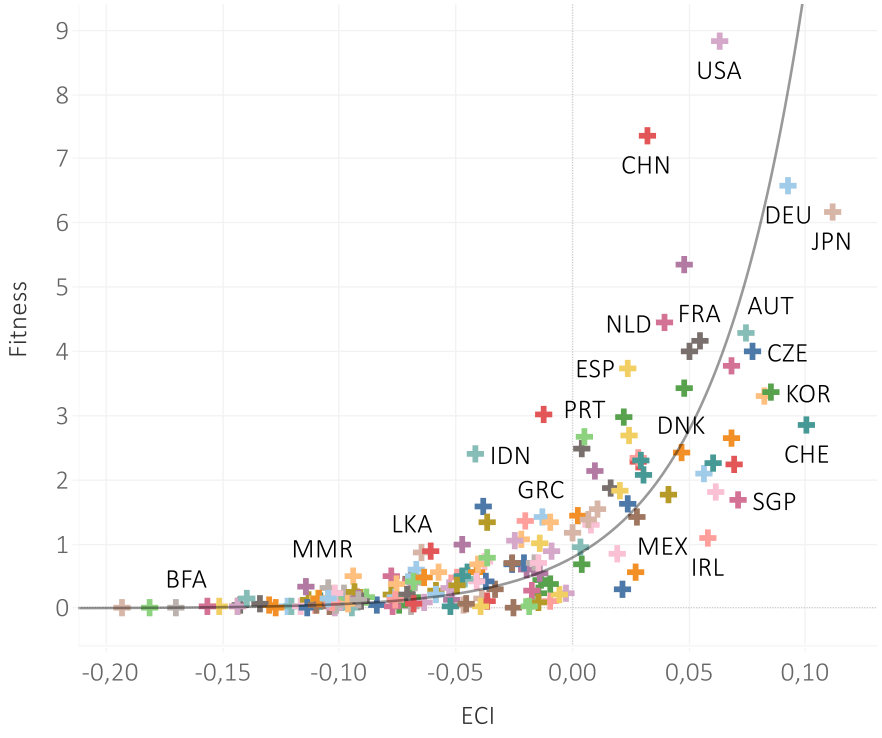


***Figure 4.5**: Scatter plot between Fitness of countries and ECI (year 2017).*

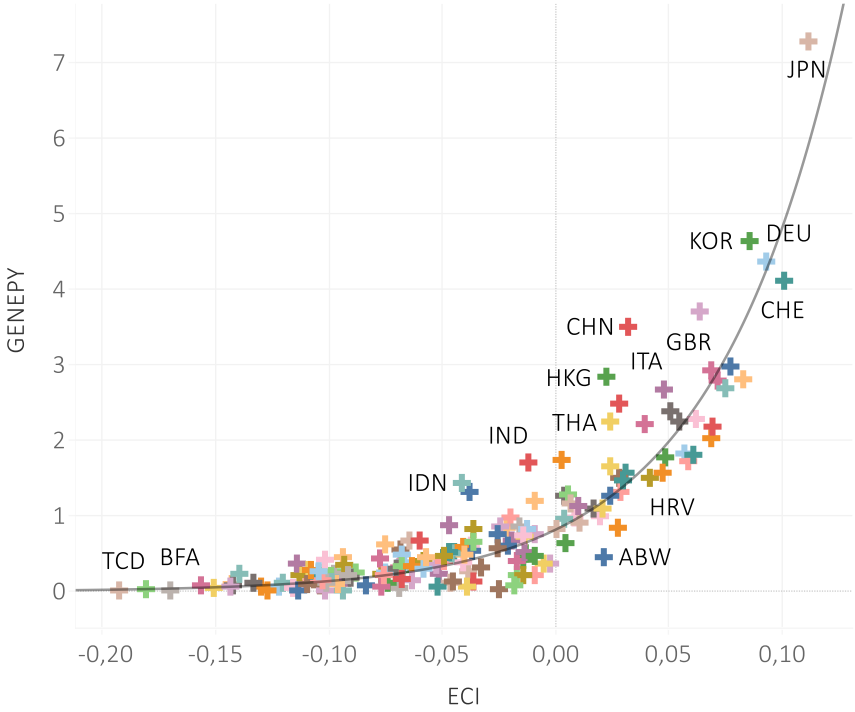


***Figure 4.6**: Scatter plot between GENEPY and ECI (year 2017).*

***Figure 4.7**: Scatter plot between GENEPY and Fitness of countries (year 2017) in logarithmic scale.*

As expected, GENEPY is well correlated with both ECI and Fitness thanks to its algebraic structure and the highest correlation occurs between GENEPY and Fitness being both non-linear methods. Therefore, it is reasonable that a higher value is obtained by correlating GENEPY and Fitness.

However, the coherence and reliability of the Reflection Method and the Fitness and Complexity algorithm has already been investigated in recent scientific works. In particular it has been highlighted that:

- the vector of the countries' complexity obtained by the algorithm of Hidalgo et al. is orthogonal to countries' diversity score (Kemp-Benedict, 2014), contrasting with the statement that complexity incorporates, in part, the diversification of the export basket of countries;
- the FC algorithm enhances economies that produce exclusive niche products, not necessarily characterized by a high level of technology (Morrison, et al., 2017), and consequently being very sensitive to the presence of noise in the dataset;
- it is still unknown which of the two metrics has greater predictive power about country's economic growth (Mariani, Vidmer, Medo, & Zhang, 2015).

According to what has been illustrated until now, we cannot confidently say what is the best metric to use to rank countries, but we can confirm that the GENEPY actually acts as a mediator towards the other two EC metrics.

The analyses done until now have concerned the comparison of available Economic Complexity algorithms by using binary incidence matrices as input. Therefore, it would be interesting to observe how the employment of matrices alternative to the binary one influences the final result. Thus, in this section we also analyse the response of the EC metrics by using different input matrices. We compared the results achieved with the binary matrix with the ones obtained by the RCA matrix and the USD matrix (described in *paragraph 3.1.1*). Results of this analysis are shown in *Table 4.2* and from *Figure 4.8* to *Figure 4.10,* in which: the suffix "bin" indicates the results achieved by using the binary matrix; the suffix "RCA" indicates the results obtained by adopting the RCA matrix; the suffix "USD$" refers to the results gained by using the USD matrix.

Y	X	Pearson's correlation coefficient
ECI_bin	ECI_RCA	-0.11
ECI_bin	ECI_USD$	0.489
Fitness_bin	Fitness_RCA	0.894
Fitness_bin	Fitness_USD$	0.750
GENEPY_bin	GENEPY_RCA	0.975
GENEPY_bin	GENEPY_USD$	0.847

***Table 4.2**: Pearson correlation coefficient between Complexity of countries calculated by changing the input matrix. Results refer to year 2017.*

As shown in Table 4.2, among the three metrics GENEPY is surely the most stable and robust algorithm to the variation of the input matrix. This is a considerable point in favour of GENEPY, as it demonstrates its applicability to weighted networks and increase its possible application outside the economic sphere. However, also Fitness and Complexity algorithm does not undergo excessive variations by changing the input matrix. A very different situation occurs when adopting the algorithm of Hidalgo and Hausmann, in fact in this case results are completely different, limiting its reliability to the use of binary matrix as input.

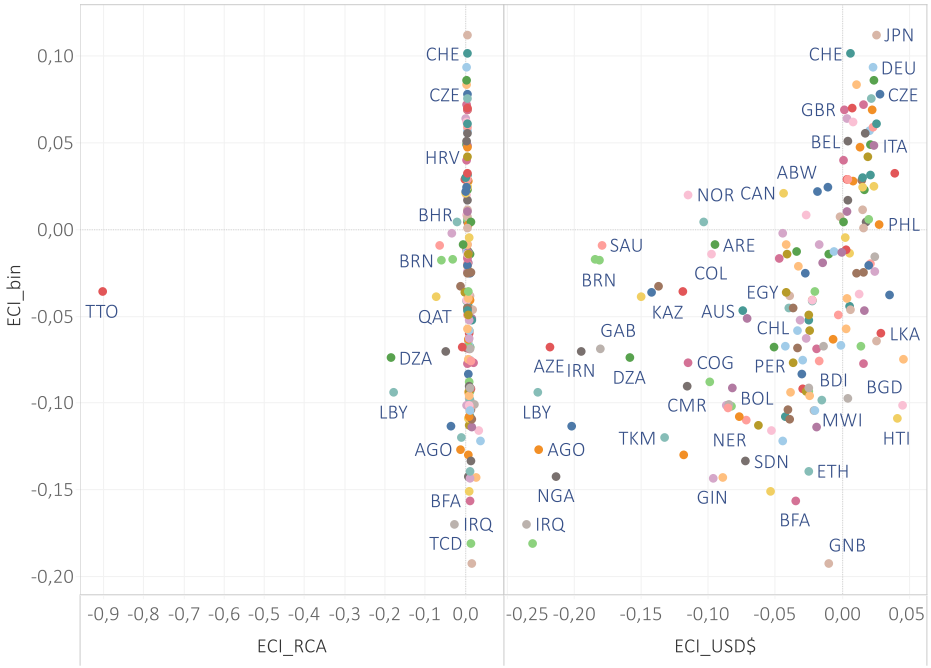


***Figure 4.8**: ECI of countries by changing input matrix (year 2017).*

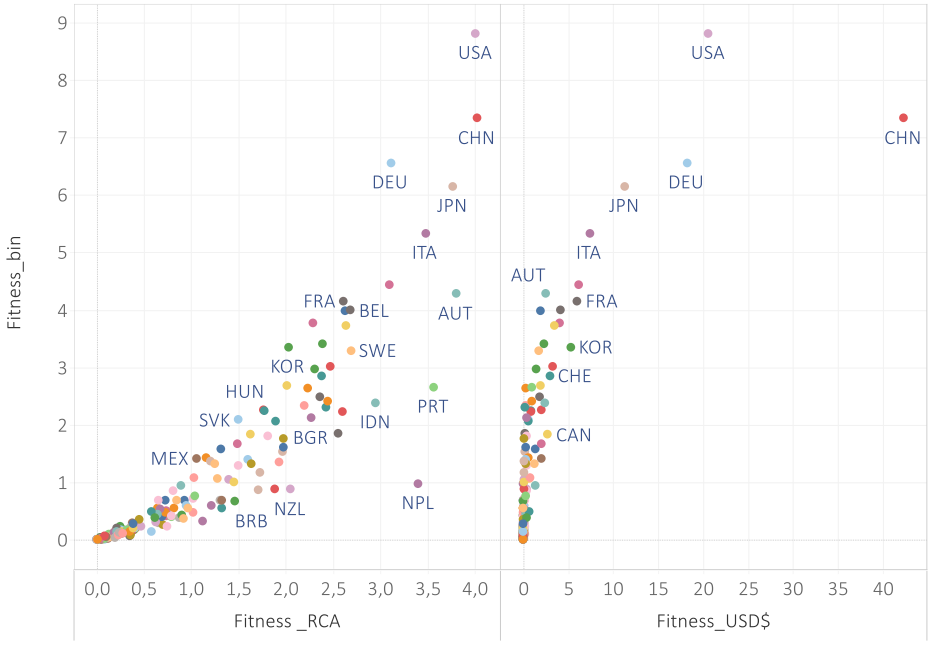


***Figure 4.9**: Fitness of countries by changing input matrix (year 2017).*

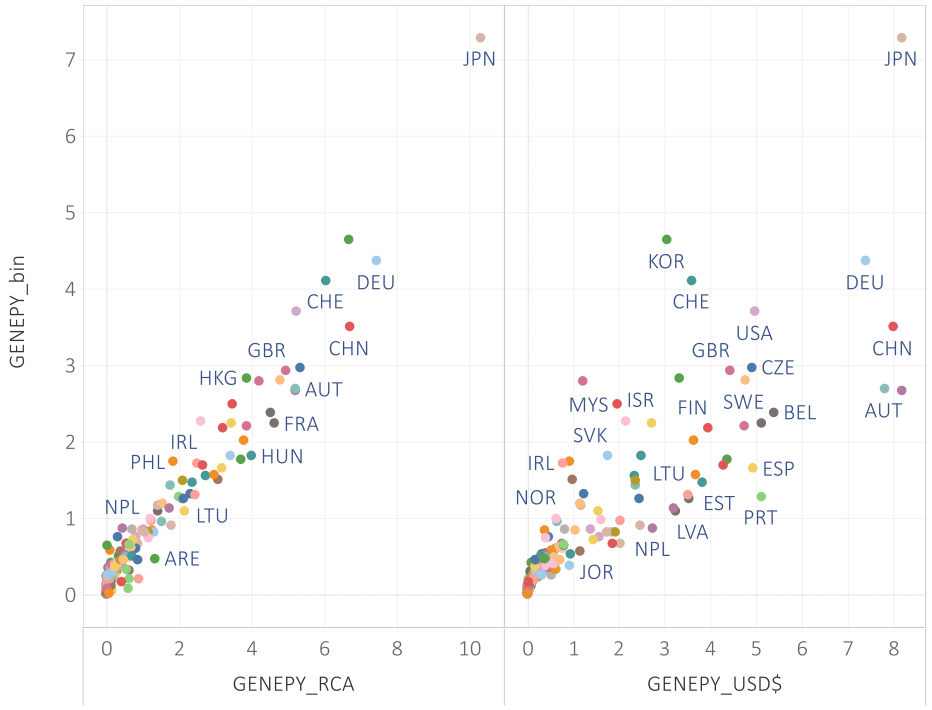


***Figure 4.10**: GENEPY of countries by changing input matrix (year 2017).*

4.2 Complexity of products

Most of the papers on Economic Complexity are focused on the reliability of the results looking mainly at countries' competitiveness, thus we dedicate this section to the exploration of products' complexity. At the 4-digit level there are more or less 1200 products, about ten times the number of countries, and consequently the products' ranking evolution in time is characterized by high variability because every year new products are introduced into the market and others are excluded. Differently, the countries in the global trade network are almost always the same and so we have been able to take information from the evolution in time of the countries' complexity ranking.

In Table 4.3 we show the correlation between results about product's complexity obtained by the three different metrics for the year 2017. The highest correlations coefficients are obtained by relating the complexity of products computed by GENEPY to the results provided by the other two metrics. Moreover, the value 0.21 concerning correlation between Quality of products (FC) and Product Complexity Index (MR) highlights the existing conceptual and algebraic difference between the two metrics.

X	Y	Pearson's correlation coefficient
PCI	$Quality_p$	0.211722732
PCI	$GENEPY_p$	0.678268443
$Quality_p$	$GENEPY_p$	0.67150672

***Table 4.3**: Pearson correlation coefficient between Complexity of products calculated by the three different EC metrics for the year 2017.*

For the same year we also show the scatter plots (*Figure 4.11*), regarding results obtained by the different EC metrics, in order to have a graphic feedback of what is shown in *Table 4.3*.

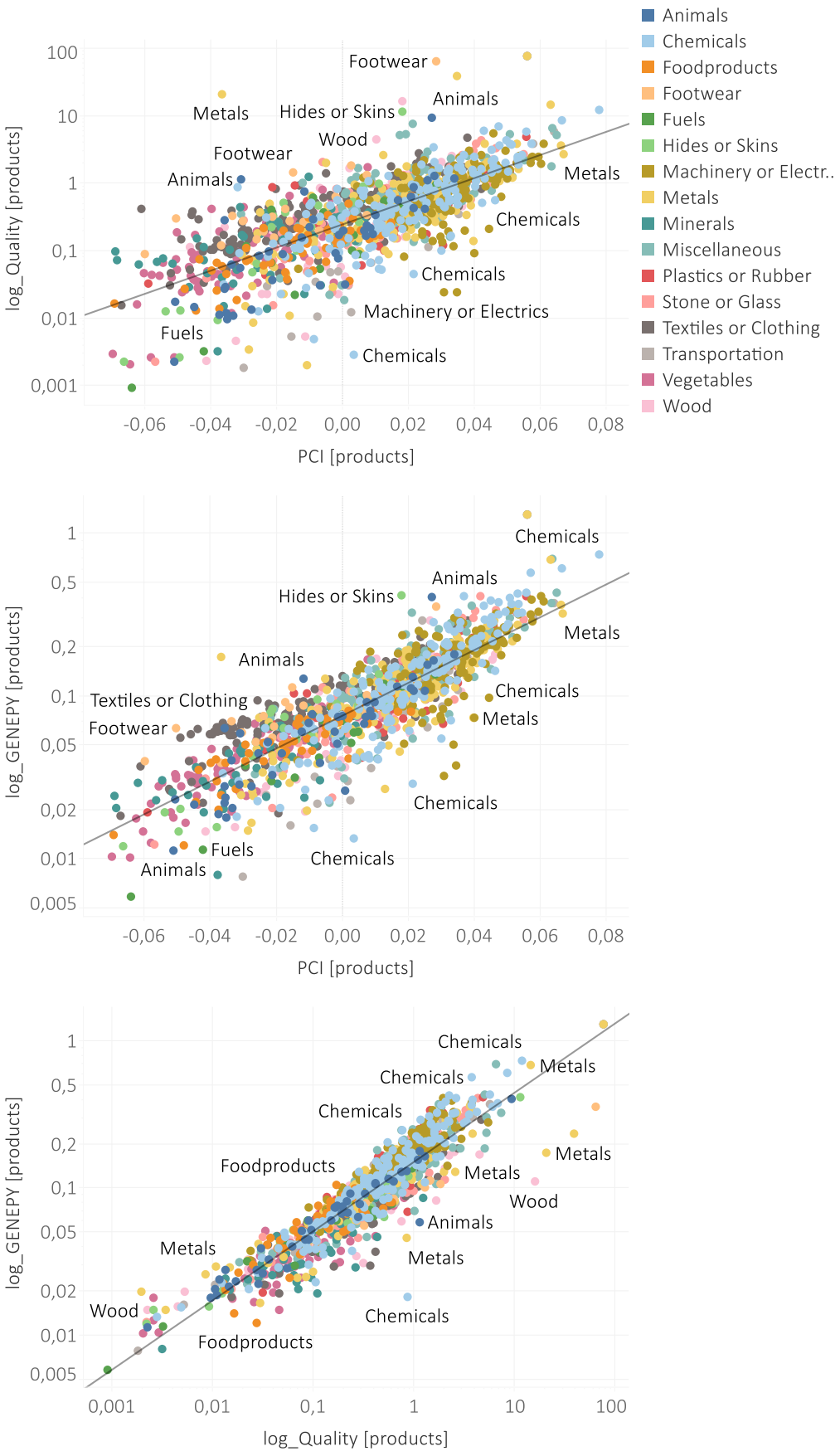


***Figure 4.11**: Comparison of products' complexities.*

In *Chapter 2* we defined *Ubiquity* as the number of countries exporting a certain product. On one hand, products exported by many countries, and therefore possessing high Ubiquity value, are probably characterized by a low level of complexity but not necessarily; on the other hand, we cannot certainly say that products exported by few countries are surely complex, even If they are exported by industrialized ones.

As shown in *Figure 4.12*, the Fitness and Complexity algorithm assigns an exceptional advantage to products characterized by extremely low ubiquity. The mentioned Figure refers to year 2017 but the same behaviour can be observed for other years. It has already been demonstrated that the FC algorithm is unstable to the presence of small disturbances in the network, specifically the computed complexity often assigns particular advantage to products exported by a very low number of countries (Morrison, et al., 2017). Thus, a detailed analysis of products' ranking is necessary to understand whether this low ubiquity products are actually distinguished by such high complexity.

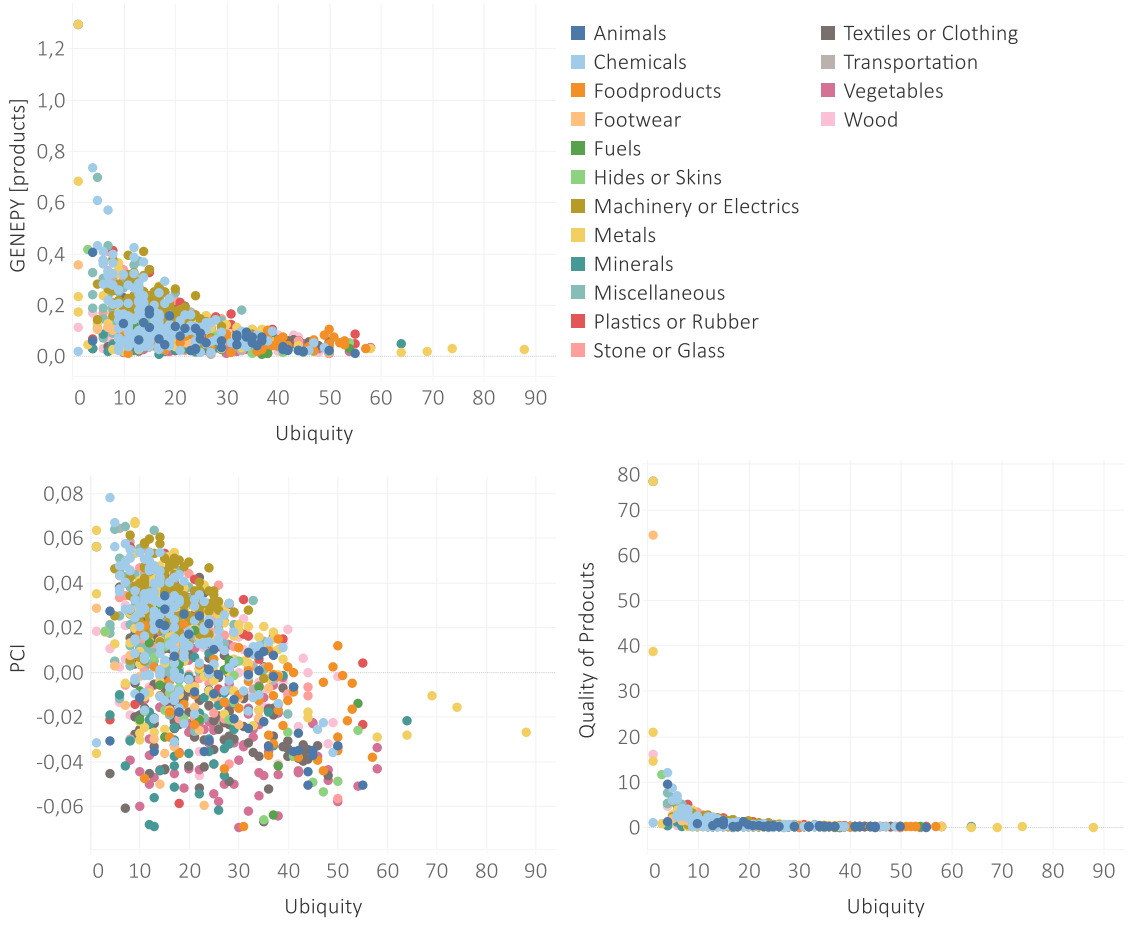


***Figure 4.12**: Complexity vs Ubiquity (year 2017).*

A first very interesting visualization concerns the results obtained for the year 1995 (*Figure 4.13*).

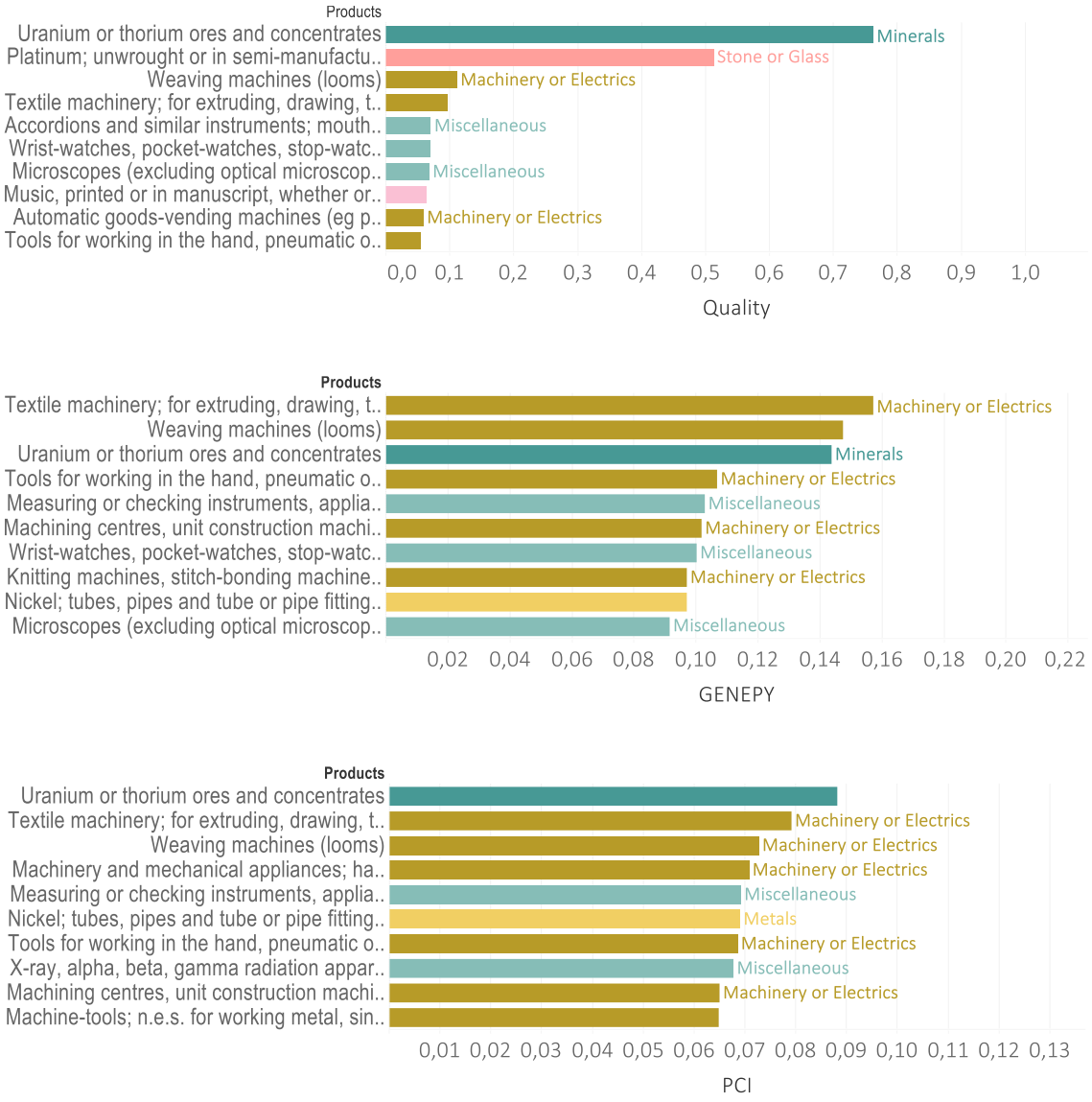


***Figure 4.13**: Top ten complex products (year 1995 – binary matrix).*

In *Figure 4.13* we exhibit the top ten complex products, ranked by the three EC metrics, for year 1995. One unexpected result is the behaviour of the item 2612 – *Uranium or thorium ores and concentrates*: this mineral, whose export depends more on natural availability than on level of actual industrial sophistication required, is at the top of the products' complexity ranking for year 1995. All the three metrics seems to recognise the complexity level of this product and in particular the FC method gives it an exceptional advantage over other products. This happens because Germany, which ranks among the top 3 complex countries during 1995, held 100% of the exports of this item, thus constituting a niche market of only

1.01 thousand dollars (The Observatory of Economic Complexity, available online, 2019) and confirming the unpleasant behaviour of the FC algorithm towards markets of this type. However, the other algorithms also overestimate the complexity of the item 2612. This anomaly would be caused by an error in the export data of 1995. Investigating on this issue, we found that an area in the eastern part of Germany, called Erzgebirge area, was a considerable source of uranium for Soviet nuclear programs between 1945 and 1989. During this time, the ex-German Democratic Republic was one of the largest producer of uranium, but the production has been stopped in 1991 after the German reunification in 1989 (Menirath, Schneider, & Menirath, 2001). Indeed, since 1991 mitigation activities has taken place in order to limit and extinguish the adverse effects of uranium mining to inhabitants.

So, basing on Germany's historical background regarding uranium production, is seems unlikely that the country holds 100% of the uranium export in 1995. This unexpected data may be present for three possible reasons:

- Uranium exported by Germany in 1995 was part of the uranium already mined when uranium mines were still open;
- It is a re-export scenario where uranium has been imported from another country before being exported by Germany;
- Presence of mistake in export data of 1995.

Whatever may be the reason, the question is "how can we avoid that a low complexity product, belonging to a niche market, is not classified as such by EC algorithms?". A possible solution may be using other input matrices. Then, in *Figure 4.14* and *4.15* we show the top ten complex products for 1995 using RCA matrix and USD matrix respectively. We do not report results concerning the Method of Reflections because we have already proved, in the results' section about countries' complexity, that when this method is applied to matrices different from the binary one, the result is unreliable. The use of the RCA matrix produces, on the one hand, a positive effect in the FC algorithm as it reduces the difference in complexity between the item 2612 and the other products; on the other hand it produces negative effects in the results of GENEPY because it brings *Uranium or thorium ores and concentrates* to the top of the complexity ranking (*Figure 4.14*). The employment of the USD matrix can instead be considered uniquely not recommended, since between the Quality (FC) of *Uranium or thorium ores and concentrates* and the underlying products (*Figure 4.15*) results a difference of four order of magnitude at least.

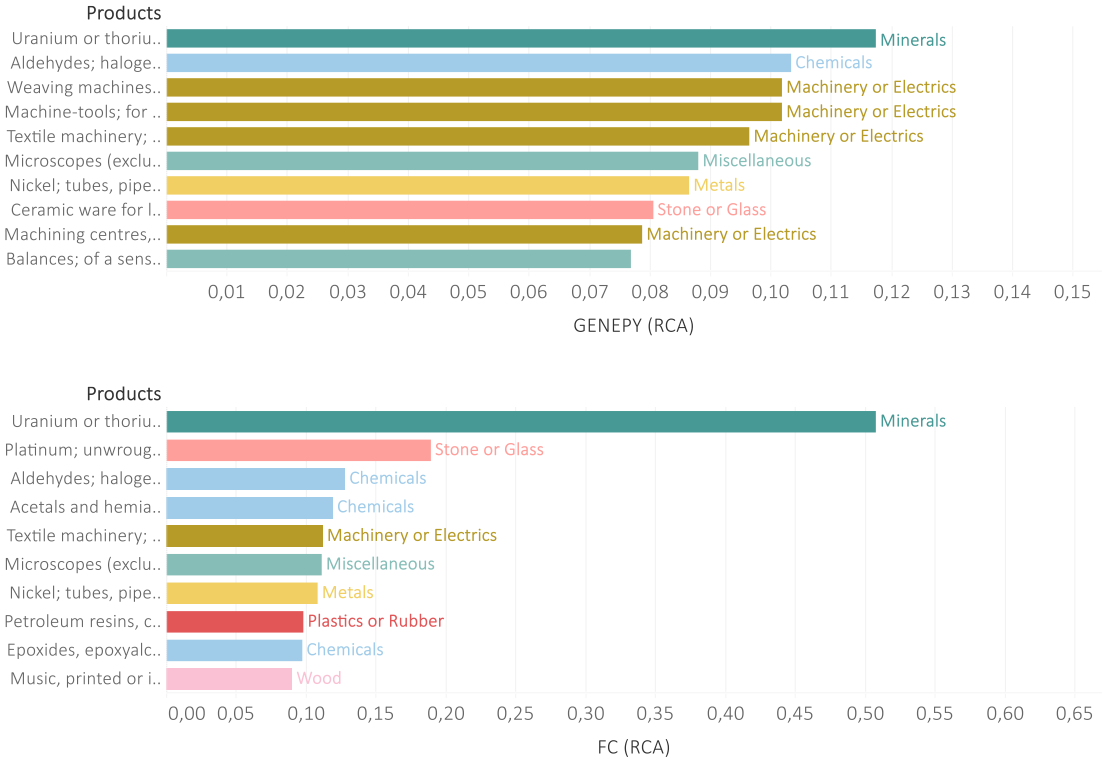


***Figure 4.14**: Top ten products for year 1995 (RCA matrix).*

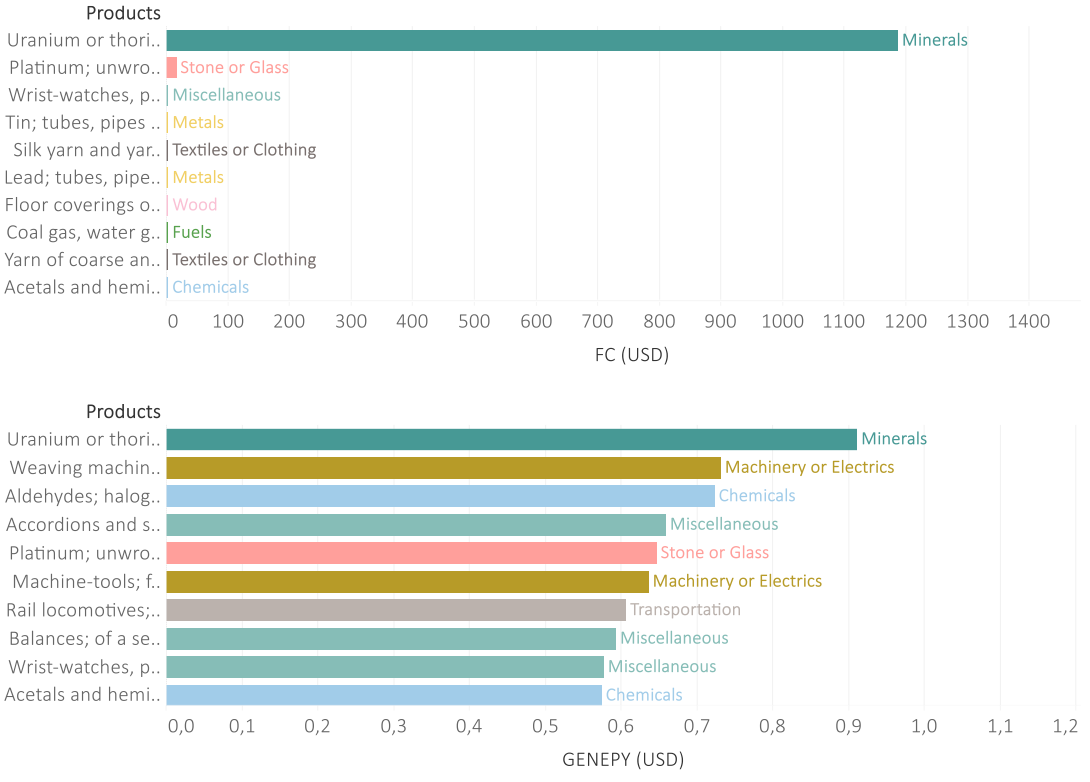


***Figure 4.15**: Top ten products for year 1995 (USD matrix).*

Item 2612 is a temporary unique product, indeed after 1995 it disappears from the market and even if it is present, it is characterized by low complexity. However, the case of *Uranium or thorium ores and concentrates* is not the only one, but similar phenomena can also be observed in other years. In *Figure 4.16* we show the top ten complex products, ranked by the three EC metrics, for year 2014 and computed using binary matrix as input.

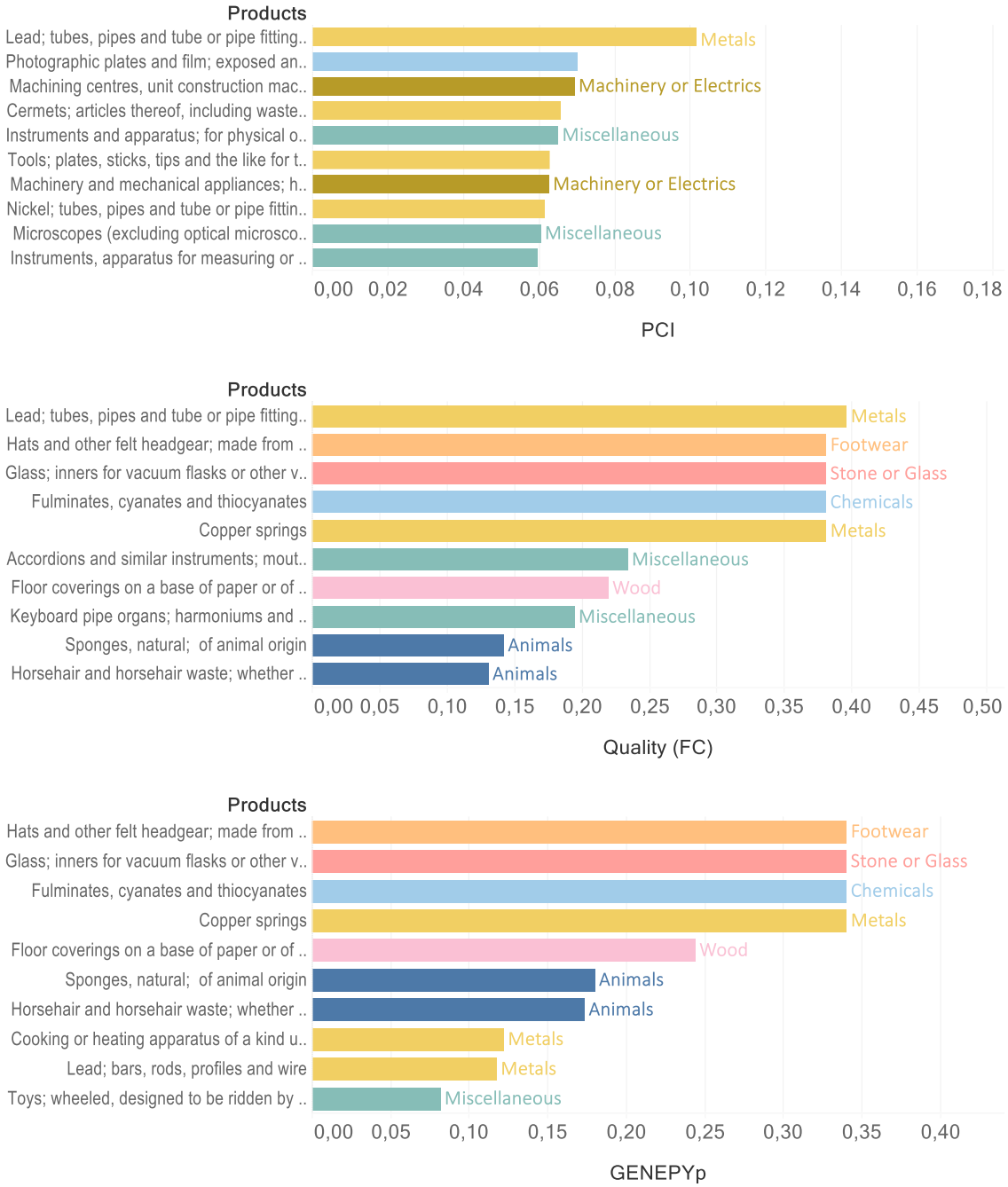


***Figure 4.16**: Top ten complex products for year 2014 (computed using binary matrix as input).*

The rankings obtained by using 2014 trade data show an apparently different situation but driven by same factors as 1995. For instance, the item 7805 – *Lead; tubes, pipes and tube or pipe fittings* is ranked at the top complex product by MR and FC. Certainly, its position in the ranking is more reasonable than *Uranium or thorium ores and concentrates,* but hardly anyone would consider it as the most complex product. In 2014 Japan, a high complex country, held 96.6% of the exports of this item, constituting a market of 735 thousand dollars (The Observatory of Economic Complexity, available online, 2019).

Therefore, about the Fitness and Complexity method, we have to agree with the observation (Morrison, et al., 2017) that complexity ranking is driven almost entirely by these temporarily unique high complex products.

4.2.1 Export basket composition

The high number of available products in the market forced us to look at the evolution of their complexity over time from a macroscopic point of view. In this section we show how the sectoral composition of the export basket of some countries changed over time, highlighting which sectors contributed more to the development of the country's economy. In order to build this visualization for each country, we divided the complexity of each product by the sum of the complexity of all exported products, then we aggregated the products complexities through their belonging sector (or category – see *Table 3.1* for more details).

We carried out this analysis for twelve countries, tying to cover different economic realities. The selected countries are: China (CHN), Brazil (BRA), Germany (DEU), United States of America (USA), Italy (ITA), India (IND), Japan (JPN), Niger (NGA), Venezuela (VEN), Philippine (PHL), Russia (RUS) and Korea (KOR). Resulting visualizations are given in *Annex A1*.

Obtained results (Annex A1) reveal again the susceptibility of the FC method to the presence of outliers, in fact the unjustified complexity of item *Uranium or thorium ores and concentrates* in 1995 translates into a much higher contribution to total complexity by the sector *Minerals* than that which characterises this sector in the following years (*Figure 4.17*). In a similar way, during years 2015 and 2016 Germany was the only relevant exporter of item 0105 - *Horsehair and horsehair waste* (ranked as the second more complex product by FC algorithm in 2015), accounting 99.5 % and 100% of the exports respectively. The temporary high complexity of this item drives the complexity of its belonging sector (*Animals*) in that years, and it has been enough to cause the *Animals* sector to contribute about 15% of the country's complexity (*Figure 4.17*).

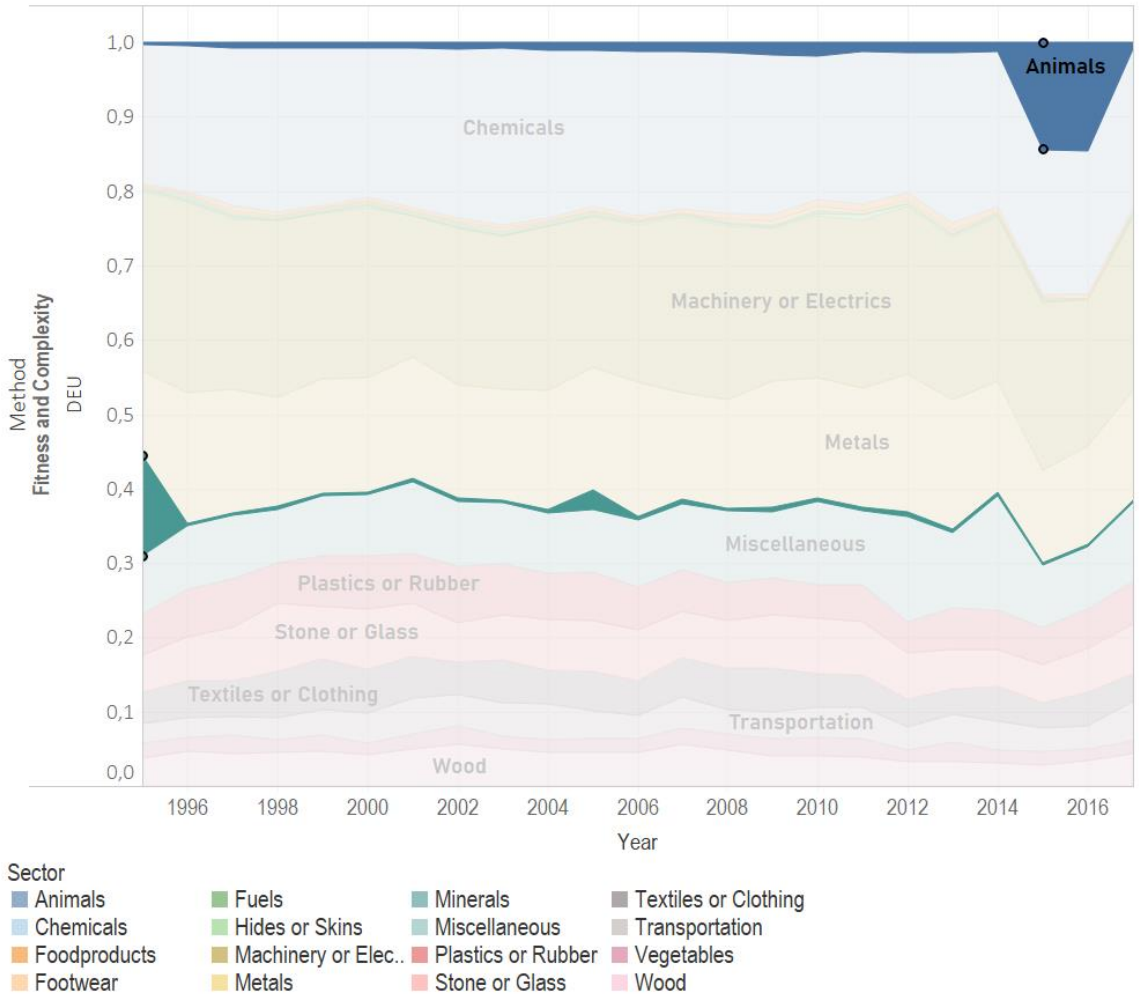


***Figure 4.17**: Export basket composition over years of DEU - Contribution of sectors Animals and Minerals to country's complexity using FC method.*

However, the three metrics agree more or less on which are the relevant sectors in the export basket of a given country. Developed countries show similar export basket composition and are characterized by a largely more stable trend, suggesting the presence of investments politics and long-term economic choices. The less industrialized countries, on the other hand, are identified by irregular composition of exports over time, typical of a subsistence economy. Perfect example of this two realities are USA and Niger (NGA) (See *Figure 4.18*).

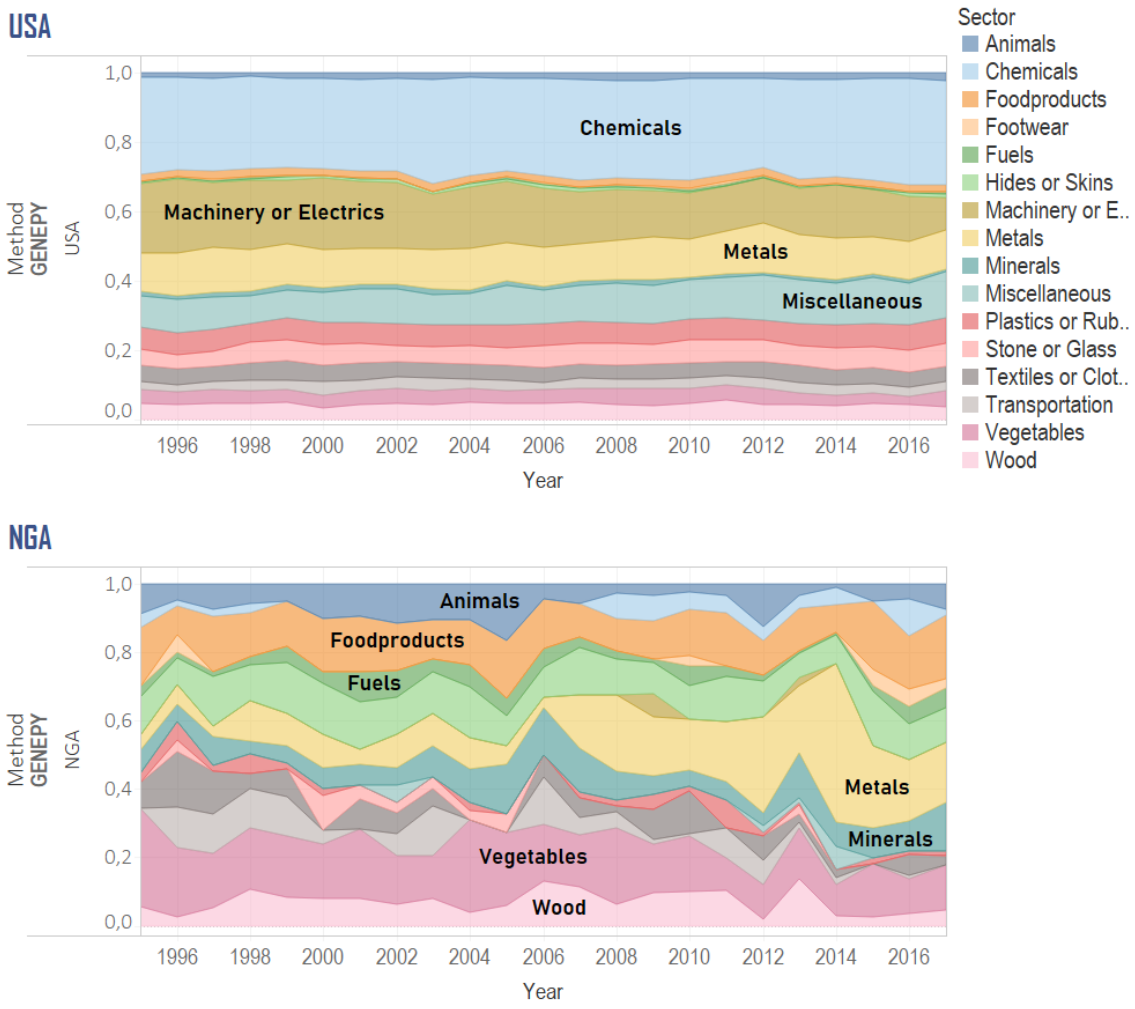


***Figure 4.18**: Export basket composition of USA and NGA using GENEPY.*

Of course, USA and Niger are two extreme economic realities, since some countries like Venezuela or Brazil are characterized by an intermediate scenario (see *Annex A1*).

4.3 Causation between GDP and Complexity

To investigate causality in the relationship between GDP per capita and EC indices, we follow the Convergent Cross Mapping method described in the previous Chapter (Materials and Methods). Applying the method proposed by Sugihara (Sugihara, et al., 2012) to our case, we have to deal with two different problems (Vinci & Benzi, 2018):

- First, the data availability: in our case we have a time series whose length is L=23, covering a time frame of 23 years (from 1995 to 2017) for each country. This is a sever limitation considering that the method has been designed for ecological applications which include sever amount of data instead;
- The second limitation is the value of the embedding dimension E: we set E=2 because we are forced to assume that Complexity (ECI, Fitness or GENEPY) and GDP are the only two relevant variables in describing economic competitiveness of countries (Vinci & Benzi, 2018).

In *Table 4.4* we show the time series of Complexity and GDP per capita concerning China.

year	Fitness	ECI	GENEPY	GDPpc (USD)
1995	3.33875	-0.01301	1.7723	609.6567
1996	3.4404	-0.01459	1.587	709.4138
1997	3.2494	-0.02074	1.4003	781.7442
1998	3.6583	-0.01523	1.6232	828.5805
1999	3.6513	-0.01335	1.7074	873.2871
2000	3.7794	-0.01441	1.778	959.3725
2001	3.8718	-0.01314	1.6581	1053.108
2002	3.9405	-0.00855	1.644	1148.508
2003	4.1675	-0.00279	2.0862	1288.643
2004	4.1691	-0.00171	2.0368	1508.668
2005	4.8399	0.00916	2.3657	1753.418
2006	4.7804	0.012007	2.3231	2099.229
2007	4.8702	0.016508	2.1627	2693.97
2008	5.078	0.02272	2.2857	3468.304
2009	4.8758	0.018636	2.127	3832.236
2010	5.4507	0.023089	2.5507	4550.454
2011	7.1659	0.024865	3.4356	5618.132
2012	6.2361	0.029023	3.2295	6316.919
2013	7.0438	0.031986	3.6838	7050.646
2014	7.8368	0.036665	3.9949	7651.366
2015	8.1654	0.033905	3.835	8033.388
2016	6.7591	0.033959	3.6127	8078.79
2017	7.345	0.032235	3.509	8759.042

***Table 4.4**: China's time series of Complexity and GDP per capita.*

A graphical representation of the time series evolution in time is given in *Figure 4.19*. In order to be able to compare the evolution in time of Complexity and GDP, the series have been normalized by dividing each value by the Frobenius norm of the series to which it belongs.

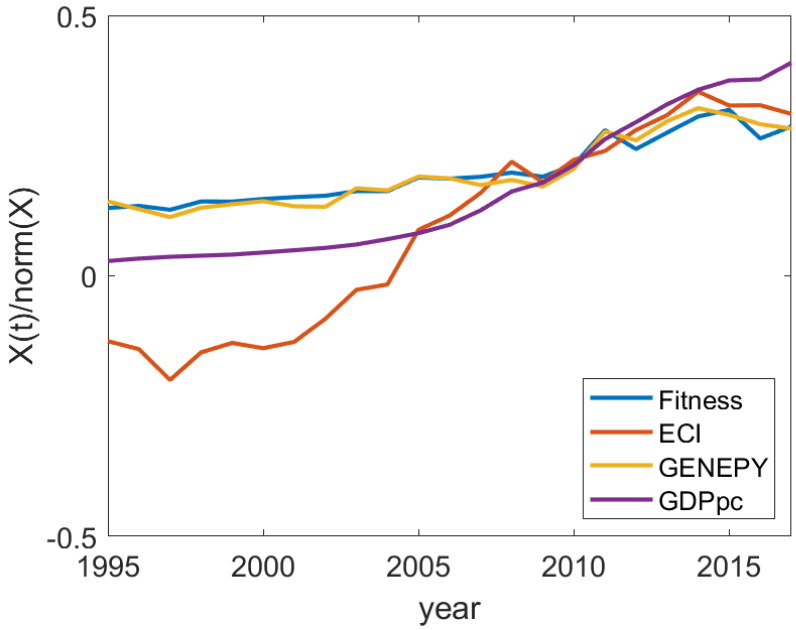


Figure 4.19*: Evolution in time of Economic Indices. Each time series is normalized dividing each value by its Frobenius' norm.*

According to Sugihara, when computing correlations between the original time series and the estimated one through another time series belonging to the same dynamic system, at increasing time series length we should appreciate an increasing correlation index if the time series are casually correlated. Thus, in this thesis work we run the CCM algorithm using time series from length L=5 to L=23; since for L<5 the dataset is too much restricted and for L>23 we have no more available data. It is necessary to observe that using L<23 we have more than one time-window satisfying the requirement of length, so we averaged the values between time windows of same length in order to increase the consistency of ours results. As an example, if L=21 we have 3 possible time windows: from 1995 to 2015, from 1996 to 2016 and from 1997 to 2017; in order to increase the robustness of our analysis we averaged the CCM correlation values using all the available time windows for a given time series length. In Figure 4.20 we exhibit the box plots of the correlations obtained by running the CCM algorithm for all the possible time windows used as input, from L=5 to L=23. GENEPY and GDPpc are the employed time series, specifically the correlations regard original series of GDPpc of China and the one reconstructed through GENEPY. Of course, increasing the length of the time series reduce the number of available time windows until is equal to 1 at L=23.

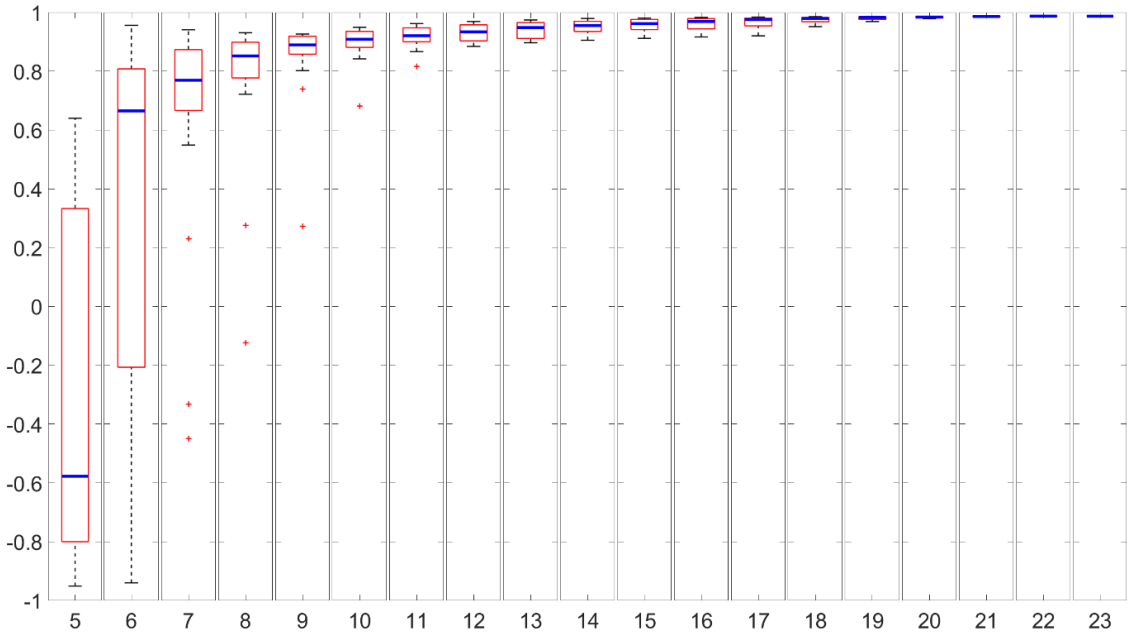


Figure 4.20: *Summary statistics of the CCM resulting from the reconstruction of the GDPpc from the GENEPY of China. The horizontal axe represents the length L and the vertical axe represents the CCM correlation.*

The CCM averaged values are plotted in *Figure 4.21*, where: the red line corresponds to the correlation obtained by reconstructing GDPpc from the country GENEPY; the blue line refers to the correlation between GENEPY and the values obtained by reconstructing GENEPY using GDPpc. As shown in *Figure 4.21* and *Table 4.5*, very high correlation values are obtained already for L>10; moreover, the plot show a bi-directional causality relationship between the two economic variables because the two curves almost cross each other. In other words, both GENEPY and GDP are able to map one-to-one each other. We consider this as an important result because confirms the ability of GENEPY (and Economic Complexity metrics) to estimate the economic competitiveness of countries.

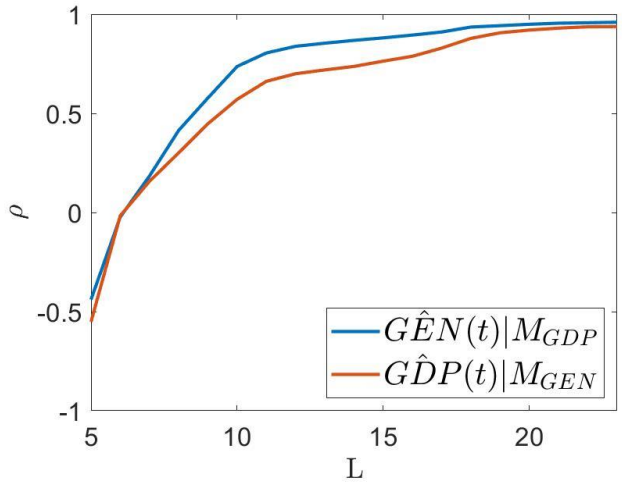


***Figure 4.21**: China: CCM between GENEPY and GDPpc. The quantity ρ-GEN/M_{GDP} (blue line) refers to the correlation obtained by reconstructing GENEPY from GDP. The quantity ρ-GDP/M_{GEN} (red line) refers to the correlation obtained by reconstructing GDP from GENEPY.*

L	ρ -GEN/M_{GDP}	ρ -GDP/M_{GEN}
1	-	-
2	-	-
3	-	-
4	-	-
5	-0.439052324	-0.55151727
6	-0.022820157	-0.016519723
7	0.183897033	0.157105534
8	0.414262602	0.301186512
9	0.577703671	0.447670276
10	0.736840013	0.570781947
11	0.80478791	0.661685547
12	0.838313911	0.699936564
13	0.853978076	0.719241069
14	0.868248616	0.73692998
15	0.880848141	0.763423863
16	0.894918542	0.788233337
17	0.910513901	0.8292296
18	0.935434777	0.87833248
19	0.942371269	0.906532056
20	0.949347113	0.920284175
21	0.954827958	0.930029012
22	0.957027543	0.936715802
23	0.959623342	0.937746001

***Table 4.5**: CCM correlations between GENEPY and GDP per capita of China.*

Once a bidirectional causal relationship between GENEPY and GDP has been established, we theorize about how we can predict one from the other. This is probably the toughest challenge and represents one of the most relevant application of the methods proposed in *paragraph 2.2*.

In 2015 a generalization of the CCM technique have been proposed (Ye, Deyle, Gilarranz, & Sugihara, 2015); in fact the general theory of CCM, based on generalized Takens' Theorem, suggests the possibility to cross map a lagged variable *X(t+lag)* from an unlagged variable *Y(t)*, considering any reasonable value of lag, because *X(t+lag)* is basically another observable variable belonging to the same dynamic system (Ye, Deyle, Gilarranz, & Sugihara, 2015). For instance, if X causes Y, the great amount of information about *Y(t)* at present should be recovered by *X(t+lag),* where *lag* is lower than zero (Vinci & Benzi, 2018). In the same way, if Y causes X, the great amount of information about future values of X should be recovered by Y at present.

In this section we present some results achieved by applying this idea to our issues. We are mainly interested in predictability of GDP using GENEPY (or any other Economic Complexity metric), thus we consider GDP per capita at time *t+lag* (in which *negative lag* corresponds to a shift in the past and *positive lag* to a shift in the future) and country complexity at time *t.* Of course, we would like to expect higher correlation when using positive lags, suggesting the possibility to predict GDP from GENEPY. We investigate predictability from lag=-6 to lag=6. Investigating this issue, we are faced again to the choice of the time series length; in fact increasing the delay between the two time series we are forced to decrease the time series length, so we fixed the time series length according to the maximum lag value: L=23-|6|=17. Of course, this assumption implies the existence of more than one time-window satisfying lags < |6| and thus the cross correlation we show in *Figure 4.22* and *Table 4.6* correspond to averaged values like we did previously. For instance, if lag = 5 we have two match satisfying this delay: GDPpc from 2001 to 2017 and GENEPY from 1996 to 2012; GDPpc from 2000 to 2016 and GENEPY from 1995 to 2011. At the expense of expectations, results suggest that the maximum correlation corresponds to lag=-4 of the GDPpc time series, which means that GENEPY better recover information about past values of GDPpc on a time scale of 4 years. However, this result is to be taken with pliers because correlation values greater than or equal to 0.8 occur throughout the entire lag range.

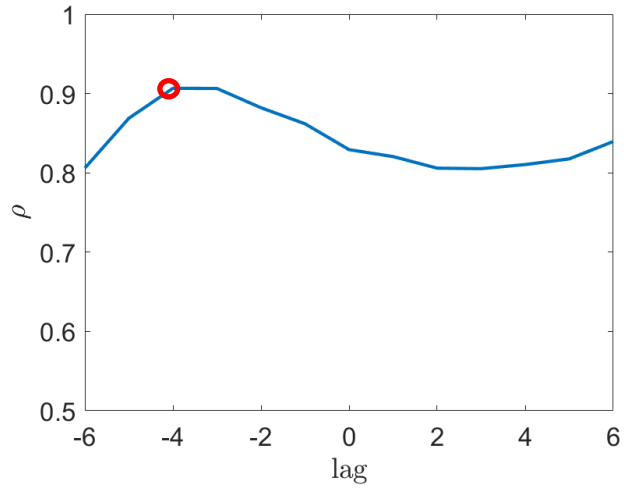


***Figure 4.22**: Correlation obtained by reconstructing GDP at time t+lag using GENEPY at time t (CHN).*

LAG	ρ - $GDP_{(t+LAG)}/GEN_{(t)}$
-6	0.806023249
-5	0.868853515
-4	0.906630768
-3	0.906385512
-2	0.882081344
-1	0.861853647
0	0.8292296
1	0.82049788
2	0.805898626
3	0.805258071
4	0.810454306
5	0.817615125
6	0.839515878

***Table 4.6**: CCM between GDPpc (t+lag) and GENEPY(t) of China.*

In order to better understand this result, it has been chosen to report also the standardised time series of the two economic indices considering fixed phase shift values between the two. Standardisation has carried out in accordance with the following formula:

$$Z = \frac{X - \mu}{\sigma}$$

Where: Z are the standardised values; X are the original values; μ and σ are the average and the standard deviation of the time series respectively. For instance, in *Figure 4.23* we present the scatter plot between GDPpc at time t-4 and GENEPY and time t.

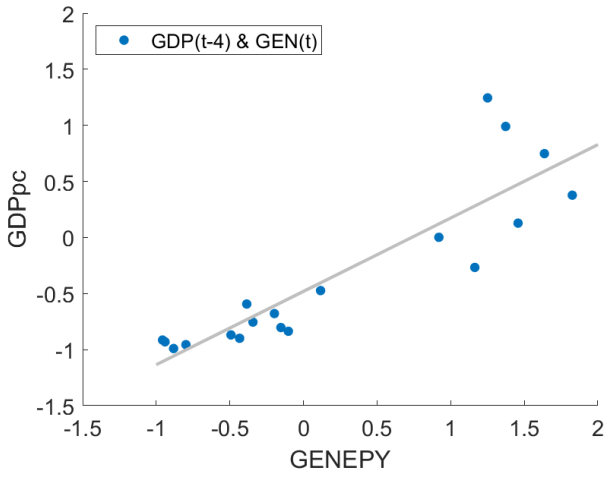


***Figure 4.23**: Scatter plot between GDP(t-4) and GENEPY(t) of CHN. Time series are normalized.*

In Annex A2 we illustrate results about causality and predictability analysis described until now, considering all the EC metrics alternatively for China, Singapore and Italy. We chose Singapore because, like China, this country increased significantly its economic competitiveness in recent years; we chose Italy instead because, differently from the other two countries, it is in a stationary economic condition since many years and thus we expect different results.

In addition, for each country we show the autocorrelation function (autocorrelation is the correlation between a time series and a delayed copy of itself) of GDPpc and of its annual increments. High autocorrelation values would explain in part the outstanding causality correlation between EC metrics and GDPpc concerning China case. Looking at figures in Annex A2, we can appreciate three different behaviours reflected by China, Italy and Singapore.

China is characterized by an extremely high autocorrelation of its GDPpc. All three EC metrics highlight the existence of a close link between GDP and Complexity, in particular the Method of Reflections. Both MR and FC method show increasing correlation for positive lags of GDP time series; specifically, the maximum value is obtained at lag=6 (result also confirmed by the relative scatter plot). GENEPY is instead characterized by a peak at lag=-4; but, unlike the other two metrics, GENEPY consist of two components (based on first two eigenvectors) and therefore assuming E=2 may not be adequate and the contribute of the two components would be analysed separately. Thus, in *Figure 4.24* we plot the cross correlation obtained reconstructing GDPpc at time *t+lag* using eigenvector associated with the largest eigenvalue

(on the left) and eigenvector associated with the second largest eigenvalue (on the right). *Figure 4.24* illustrate a different behaviour, showing higher correlation for positive lags. However, more in depth analyses are required in order to understand why GENEPY behaves differently.

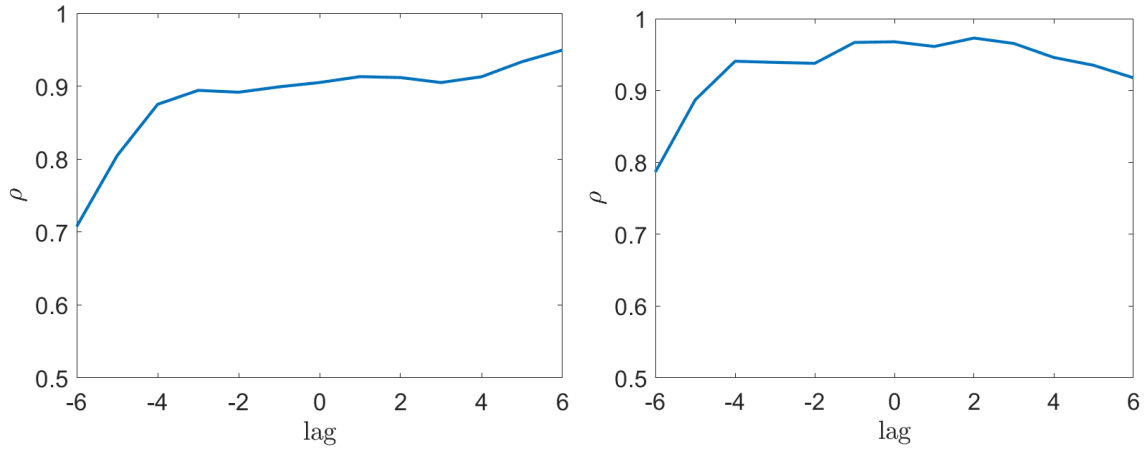


***Figure 4.24**: Cross Correlation between delayed time series of GDPpc and both first eigenvector (on the left) and second eigenvector (on the right).*

Italy and Singapore instead behave in extremely different ways. Both these case studies highlight a chaotic behaviour and correlations hardly exceed the value of 0.5, even if considering the entire time series length. Unlike Singapore, Italy is characterized by a considerable deviation between the CCM curves, thus resulting in the absence of a bidirectional causality relationship; moreover, only the correlations obtained by reconstructing GENEPY from GDP identify a growing asymptotic trend as the length of the time series increases. About Singapore instead, even if the CCM curves rise as L increase, there is no evidence of asymptotes, suggesting the possibility that dataset is too small. Looking at results concerning cross correlations of Italy in Annex A2, figures show a clear peak both at negative and positive lags when using MR and FC method. About Singapore instead, MR and GENEPY show a clear positive peak at lag equal to 6. However, data availability is our major constrain because correlated time series are too short and thus the reliability of the results and the possible range of investigated lags is limited. On the basis of what have been shown until now we can certainly state that a causal correlation between country complexity and GDP exists. On the other hand, we don't fully understand how GDP and complexity can mutually interact with each other and what is the optimal time scale to take into account within the single country.

CHAPTER 5

5 Conclusions

Improving data-driven economic theories is a crucial step to understanding the complex phenomenon of economic development. Different metrics have been proposed in order to quantify the intangible complexity of countries and products. The available Economic Complexity metrics are the Method of Reflections (Hidalgo & Hausman, 2009) and the Fitness and Complexity algorithm (Tacchella, Cristelli, Caldarelli, Gabrielli, & Pietronero, 2012). This two metrics diverge among them conceptually and thus in the provided results. A team from Politecnico di Torino have recently designed a new algorithm, called GENEPY, able to reconcile the two existing metrics and providing a unique ranking for countries and products (Sciarra, Chiarotti, Ridolfi, & Laio, 2020). Therefore, in this section we would like to resume the main results of our analysis, highlighting the main differences between EC methods and suggesting future possibilities that would help in improving the importance of data-driven economic theories.

About the ranking of countries, the differences are quite small, as confirmed by the high values of Pearson's correlation index. When considering the weighted bipartite network, and thus employing input matrices different from the binary one, the Method of Reflections provide unreliable results. On the other hand, GENEPY is the most stable method and in particular a Pearson's Index of 0.975 is obtained when correlating rankings achieved through RCA and binary matrices. The consistency of a metric even in the presence of weighted networks is certainly an important feature, especially when extending its applicability to networks different from the economic one.

We found more substantial differences in the correlations concerning the complexity of products. The Fitness and Complexity algorithm has proven to be vulnerable when a high-income country owns the totality of the exports of a niche product, overestimating its complexity. The behaviour of item 2612 – *Uranium or thorium ores and concentrates* is a

perfect example of this problem; in fact, this product drove the complexity of its whole belonging sector (Minerals) during year 1995, contributing alone to about 15% of Fitness of Germany. GENEPY and MR have more robust architecture, but still overestimate the complexity of products of this type. The problem persists when using input matrices different from the binary one, even if RCA matrix influence quite positively the FC results. The analysis of the evolution in time of the export basket composition allowed us to obtain information about the industrial sophistication of some countries, and which sectors are contributing more to their overall complexity. Developed countries are typically characterized by investments-based economies, reflecting a mostly stable export basket composition. Poor countries instead show high variability in exports, suggesting the employment of sustenance-based economy and no long-term economic politics.

Last but not least, we investigated causality and predictability between Complexity and Gross Domestic Product per capita. In order to make this analysis we followed the method proposed in (Sugihara, et al., 2012), called Convergent Cross Mapping. When applying this technique, we assumed that Complexity and GDPpc are the only two variables contributing to country economic performance. Although the restricted data availability, we have been able to show that a clear causal correlation between GDP per capita and Economic Complexity exists, especially when analysing China case study. In order to investigate predictability, we analysed the effect of time *lag* in the causal correlation. For China economy we observed that, when correlating GDP with both MR and FC, the highest correlation value occurs on time scale of about 6 years, but the plots do not show a clear peak, and therefore we cannot confidently establish what is the better lag to use. Looking at the effect of time *lag* in the causal correlation between GDP and GENEPY, we observed a clear peak at *lag=-4*, meaning that GENEPY better recover past values of GDP. However, GENEPY has two components, based on the first two eigenvectors, so the assumption E=2 may not be an adequate configuration. Indeed, as shown in *Figure 2.24* the first two eigenvectors recover better information about GDP for positive *lags*. For other countries, like Singapore and Italy, the connection GDP-Complexity is weaker but still present, and the effects of time *lag* in the causal correlations suggest that is unrealistic to find a unique prediction time scale valid for all countries and metrics. Thus, we argued that both GDP and Complexity time series are connected at different time scales in a complex dynamic system. Of course, more in-depth investigations are needed in order state the validity of the results and understand how we can use them. Furthermore, explaining the different behaviour of China, Singapore and Italy is a tough challenge to be faced in the future.

Bibliography

Abdon, A., Bacate, M., Felipe, J., & Kumar, U. (2010, September). Product Complexity and Economic Development. *Levy Economics Institute Working Paper Collection*.

Balassa, B. (1965). Trade liberalization and 'revealed' comparative advantage. *Manchester School*, 99-123.

Barabasi, A.-L. (2016). *Network Science.* Retrieved from http://networksciencebook.com/

Bar-Yam, Y. (2014). General Features of Complex Systems. *Encyclopedia of Life Suppoprt Systems*.

Callen, T. (2018, December). *International Monetary Fund*. Retrieved from https://www.imf.org/external/pubs/ft/fandd/basics/gdp.htm

CEPII. (n.d.). Retrieved from http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=37

CIA World Factbook. (n.d.). *index mundi*. Retrieved from https://www.indexmundi.com/map/?v=65&l=it

Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G., & Pietronero, L. (2013, August). Measuring the Intangibles: A Metric for the Economic Complexity of Counrtries and Products. *PLOS ONE*. doi:10.1371/journal.pone.0070726

Gaulier, G., & Zignago, S. (2009). BACI: International Trade Database at the Product-level. *MPRA*, 33. Retrieved from https://mpra.ub.uni-muenchen.de/31398/

Golub, G., & Van Loan, C. (2013). *Matrix computations.* Baltimore: The John Hopkins University Press.

Hausmann, R. (2013). *Atlas of Economic Complexity*. Retrieved from http://atlas.cid.harvard.edu/our-team

Hausmann, R. (n.d.). *Atlas of Economic Complexity*. Retrieved from http://atlas.cid.harvard.edu/our-team

Hausmann, R., & Hidalgo, C. (2010, October). Country diversification, product ubiquity, and economic divergence. *Working Papers (Center for International Development at Harvard University)*. Retrieved from http://atlas.cid.harvard.edu/publication-archive

Hausmann, R., Hwang, J., & Rodrik, D. (2007). What you export matters. *J. economic growth*.

Hidalgo, C., & Hausman, R. (2009, June). The buildig blocks of economic complexity. *PNAS*.

Huke, J. (2006). *Embedding Nonlinear Dynamical Systems: A Guide to Takens' Theorem.* The University of Manchester, School of Mathematics. Manchester: Manchester Institute for Mathematical Sciences . Retrieved from http://eprints.maths.manchester.ac.uk/

International Monetary Fund. (n.d.). *Report for Selected Countries and Subjects*. Retrieved from imf.org: https://www.imf.org/external/pubs/ft/weo/2013/01/weodata/weorept.aspx?sy=1980&ey=2018&sort=country&ds=.&br=1&pr1.x=40&pr1.y=0&c=924&s=NGDP_RPCH%2CPPPPC&grp=0&a=

Kemp-Benedict, E. (2014, Dicembre). An interpretation and critique of the Method of Reflections. *Munich Personal RePRc Arichive*. Retrieved from https://mpra.ub.uni-muenchen.de/60705/

Krakovskà, A., Jakubìk, J., Budacova, H., & Holecyova, M. (2016). *Causality studied in reconstructed state space. Examples of uni-directional connected chaotic systems.* Institute od Measurement Science, Slovak Academy od Sciences, Bratislava.

Kuhn, R. L. (2013, June 4). Xi Jinping'd Chinese Dream. *The New York Times*.

Liang, Q., & Teng, J.-Z. (2005). Financial development and economic growth: Evidence from China. *ELSEVIER*. Retrieved from https://doi.org/10.1016/j.chieco.2005.09.003

Mariani, M. S., Vidmer, A., Medo, M., & Zhang, Y.-C. (2015). Measuring economic complexity of countries and products: which metric to use? *The European Physical Journal B*. doi:10.1140/epjb/e2015-60298-7

McCracken, J. M., & Weige, R. S. (2014). *Convergent Cross-Mapping and Pairwise Asymmetric Inference.* George Mason University, School of Physics, Astronomy and Computational Science. Retrieved from https://arxiv.org/abs/1407.5696v1

Menirath, A., Schneider, P., & Menirath, G. (2001). Uranium ores and depleted uranium in the environment, with a reference to uranium in the biosphere from the Erzgebrige/Schnes, Germany. *Journal of Environmental Radioactivity*. Retrieved from www.elsevier.com/locate/jenvrad

Morrison, G., Buldyrev, S., Imbruno, M., Arrieta, O., Rungi, A., Riccaboni, M., & Pamomolli, F. (2017, November). On Economic Complexity and the Fitness of Nations. *Scientifc Reports*. doi:10.1038/s41598-017-14603-6

Nathas, L., Oswald, F., & Nimon, K. (2012). Interpreting multiple linear regression: A guidebook of variable importance. *Pract. Assessment, Res.& Eval.*

Nayfeh, A. H., & Balachandran, B. (2008). *Applied nonlinear dynamics.* WILEY-VCH.

Negre, C. F., Morzan, U. N., Hendtrickson, H. P., Pal, R., Lisi, G. P., Loria, J., . . . Batista, V. S. (2018). Eigenvector centrality for characterization of protein allosteric pathways. *PNAS*. doi:10.1073/pnas.1810452115

Newman, M. (2006). *The mathematics of networks.* Center fo the Study of Complex Systems, University of Michigan.

Newman, M. (2010). *Networks: An introduction.* Oxford University Press.

Newmann, M. (2006). Finding Community structure in networks using the eigenvector of matrices. . *Phys.Rev.E*, 74.

Pietronero, L., Cristelli, M., & Tacchella, A. (2013, April). New Metrics for Economic Complexity: Measuring the Intangible Growth Potential of Countries. *Conference of the Institute for New Economic Thinking*.

Sciarra, C., Chiarotti, G., Laio, F., & Ridolfi, L. (2018, October). A change of perspective in network centrality. *Scientific Reports*.

Sciarra, C., Chiarotti, G., Ridolfi, L., & Laio, F. (2020, July). Reconciling contrasting views on economic complexity. *Nat Commun*, 11. doi:https://doi.org/10.1038/s41467-020-16992-1

Servedio, V., Buttà, P., Mazzilli, D., Tacchella, A., & Pietronero, L. (2018, October). A new and stable estimation method of country economic fitness and product complexity. *Cornell University*. doi:10.3390/e20100783

Simoes, A., Landry, D., & Hidalgo, C. (n.d.). *Observatory of Economic Complexity*. (A. Simoes, Editor) Retrieved from https://oec.world/en/resources/methodology/

Sims, D. (n.d.). *China Widens Lead as World's Largest Manufacturer*. Retrieved from THOMAS.

Smith, A. (1776). *The Wealth of Nations.*

Stewart, I. (2000, August). The Lorentz attractor exists. *NATURE*.

Sugihara, G., Grenfell, B. T., & Mccreddie May, R. (1990, November 29). Distinguishing error from chaos in ecological time series. *The Royal Society*. Retrieved from https://doi.org/10.1098/rstb.1990.0195

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarity, M., & Munich, S. (2012). Detecting Causality in Complex Ecosystems. *Science*. doi:10.1126/science.1227079

Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., & Pietronero, L. (2012, October). A New Metrics of Countries' Fitness and Products' Complexity. *Sientific Reports*. doi:10.1038/srep00723

Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture Notes in Mathematics, vol 898*. doi:https://doi.org/10.1007/BFb0091924

Tu, C., Carr, J., & Suweis, S. (2016). A Data Driven Newkork Approach to Rank Countries Production Diversity and Food Specialization. *PLoS ONE*. doi:10.1371/journal.

UN International Trade Statistics. (2017). Retrieved from UN comtrade: https://unstats.un.org/unsd/tradekb/Knowledgebase/50018/Harmonized-Commodity-Description-and-Coding-Systems-HS

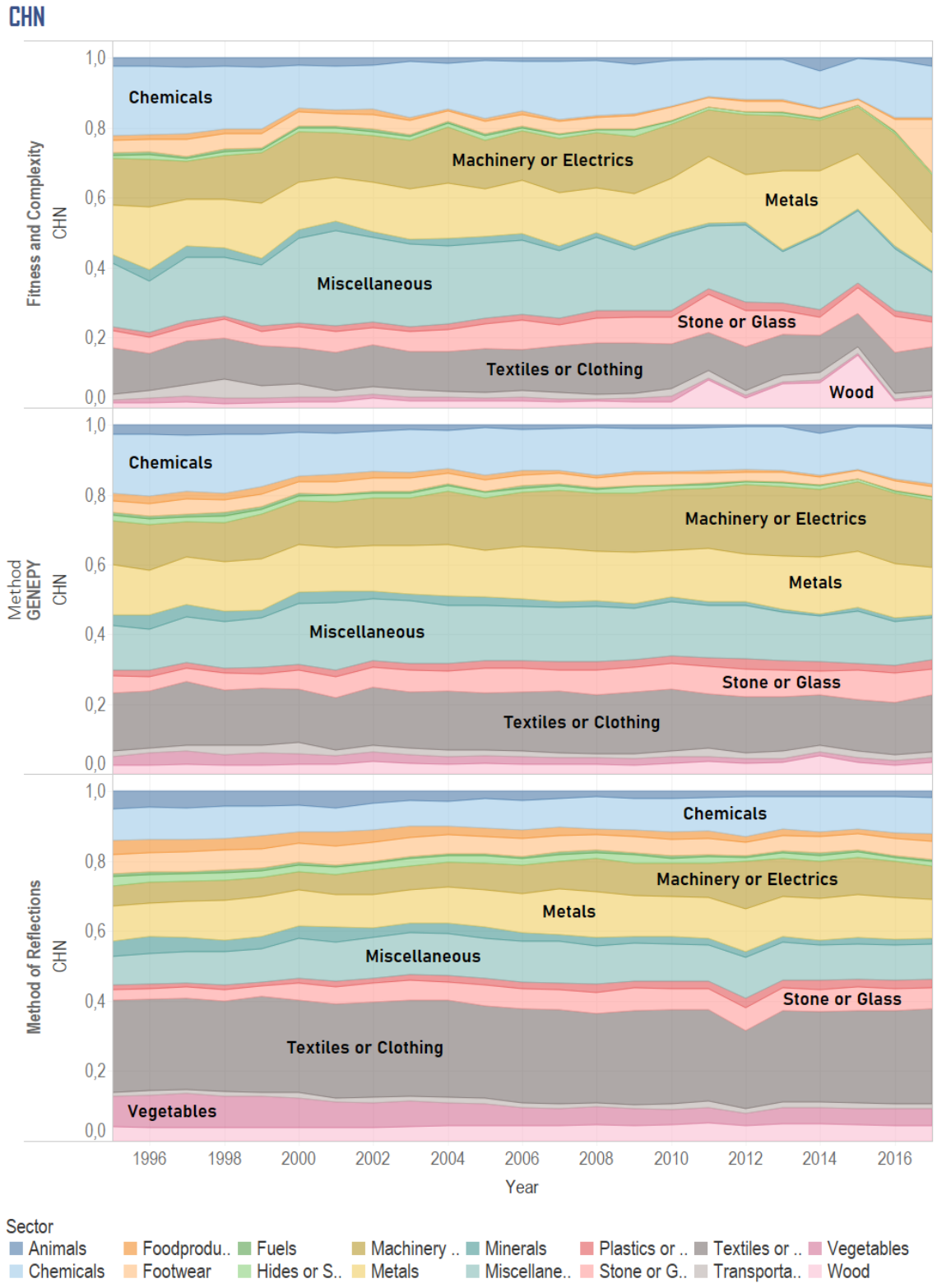
Vinci, G. V., & Benzi, R. (2018). Economic Complexity: Correlations between Gross Domestic Product and Fitness. *entropy*. doi:10.3390/e20100766

William, R., & Martinez, N. (2000). Simple rules yeald complex food webs. *Nature 404*. Retrieved from https://doi.org/10.1038/35004572

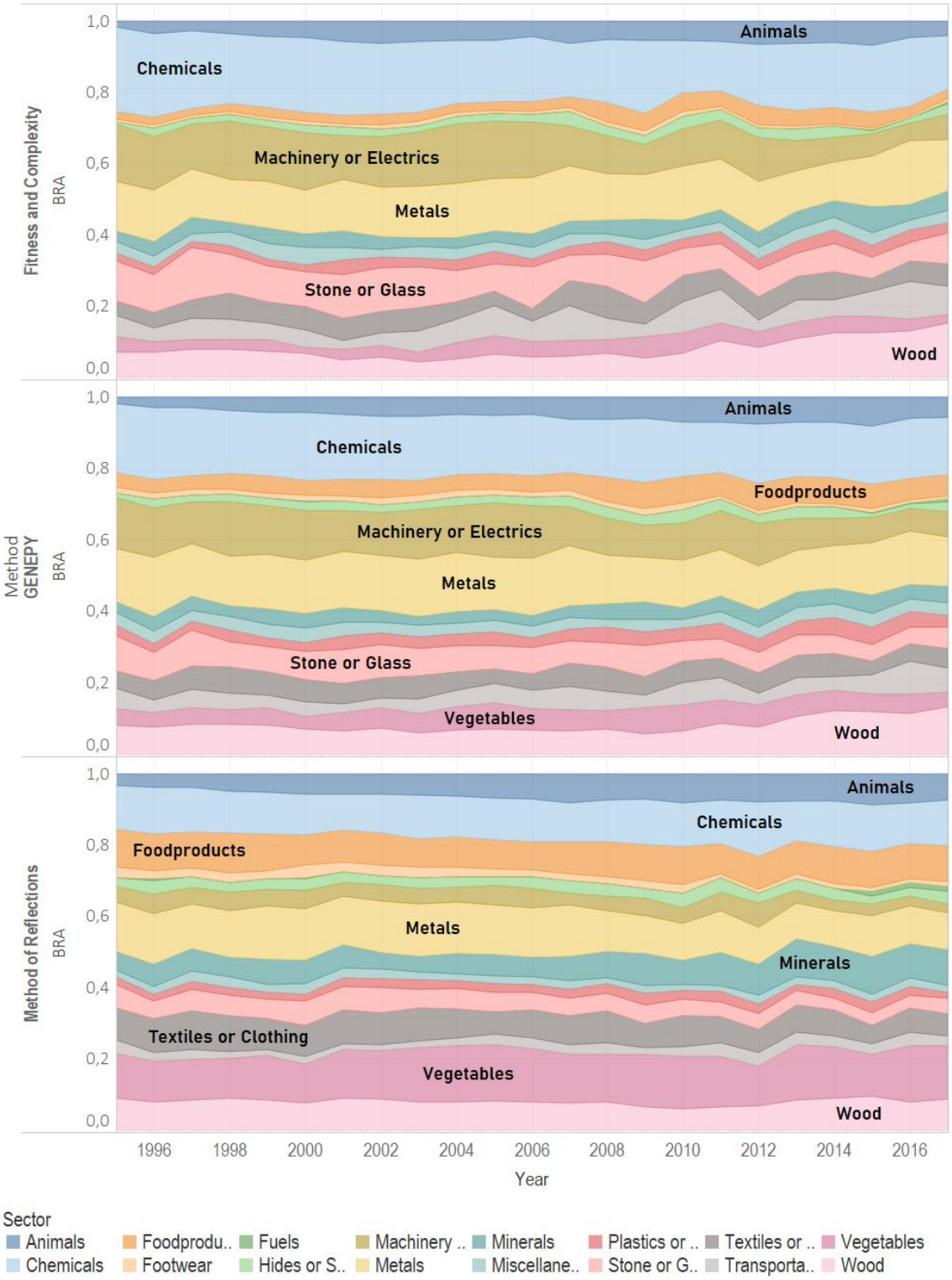
Willy, C., Neugebauer, E. A., & Gerngrob, H. (2003). The Concept of Nonlinearity in Complex Systems. *European Journal of Trauma*, 29. Retrieved from https://doi.org/10.1007/s00068-003-1248-x

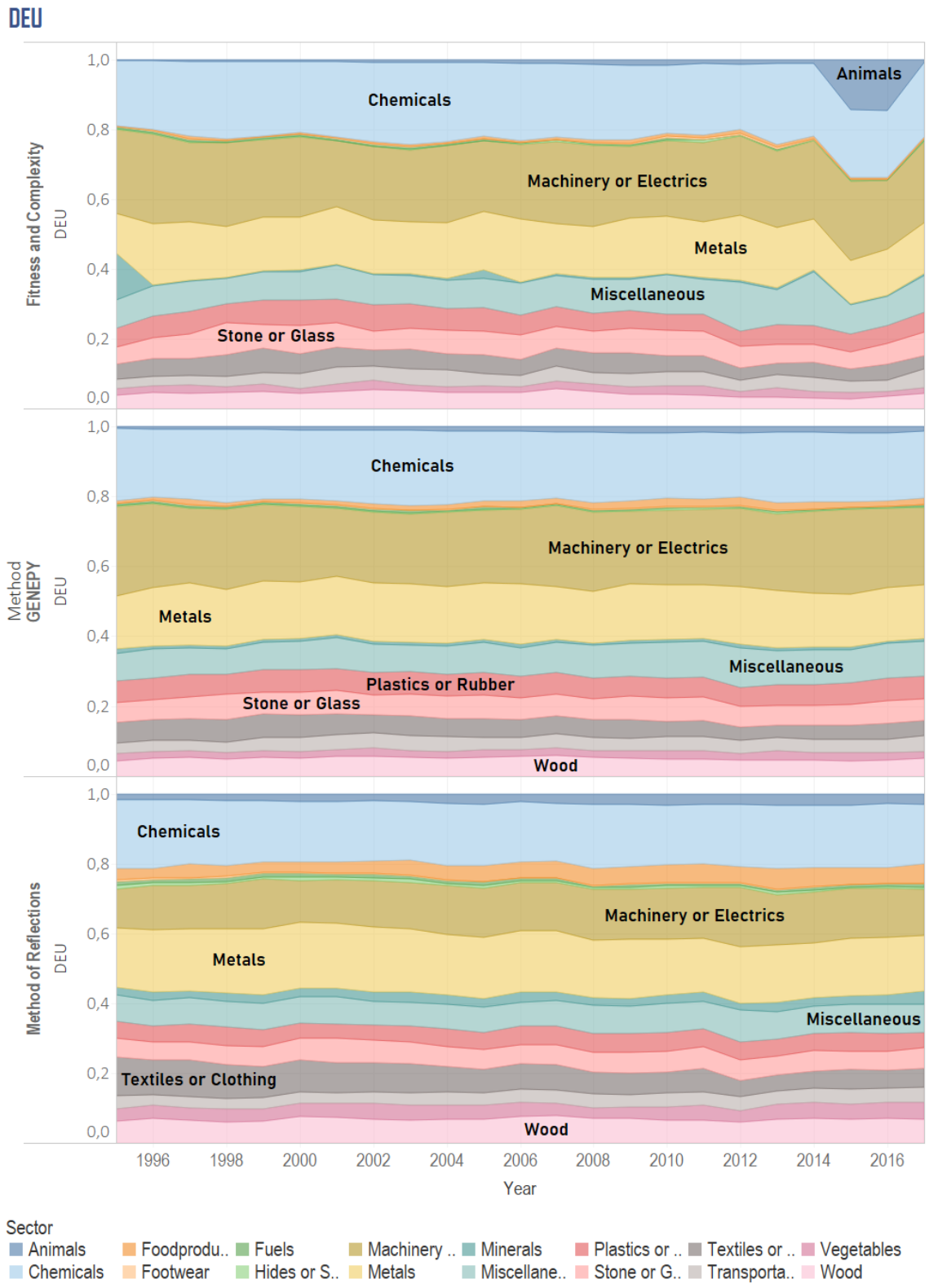
Ye, H., Deyle, E. R., Gilarranz, L. J., & Sugihara, G. (2015). Distinguishing time-delayed causal interactions using convergent cross mapping. *Scientific Reports*. doi:10.1038/srep14750

Annex A1

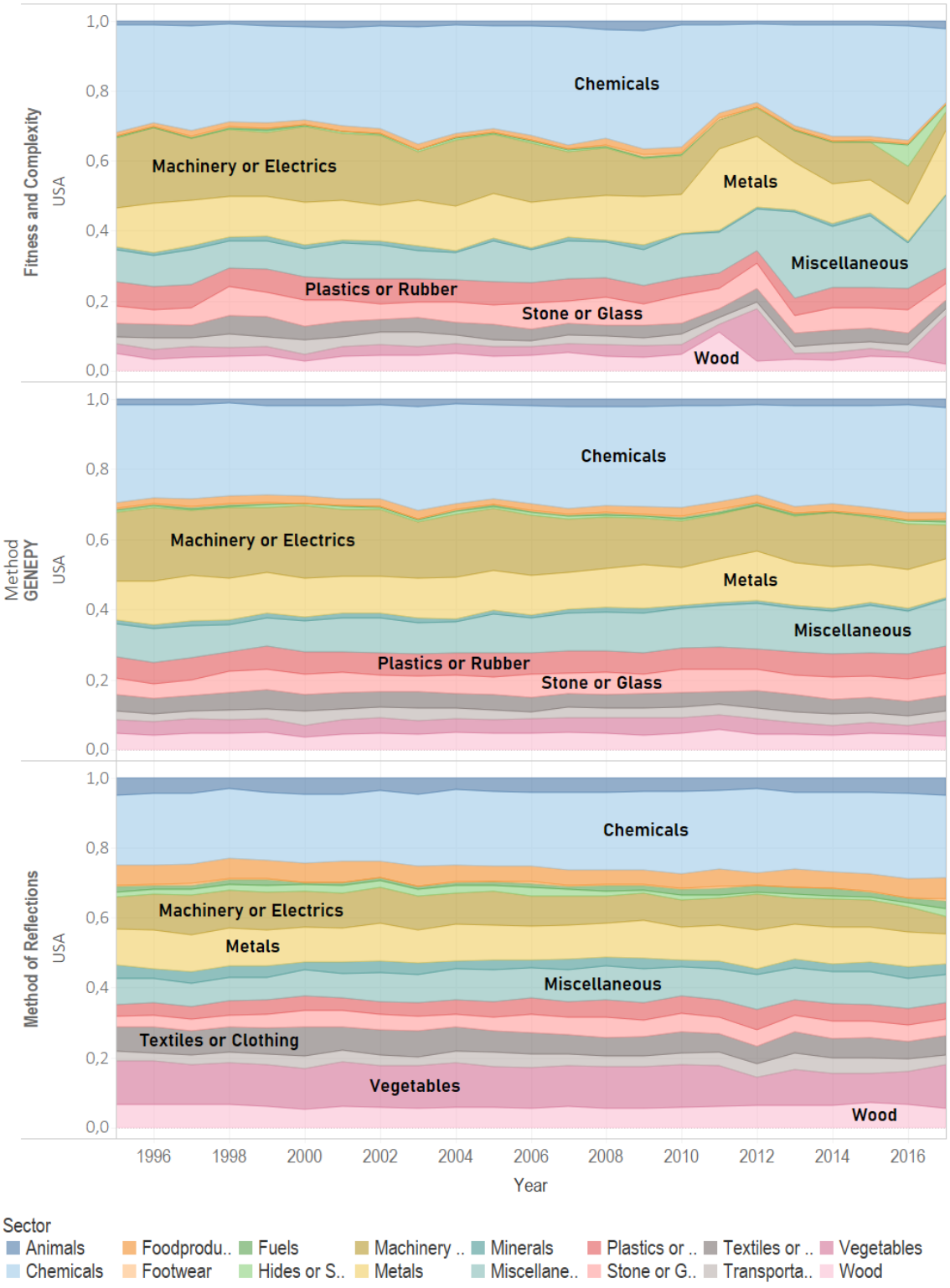

CHN
Fitness and Complexity CHN
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Stone or Glass
Textiles or Clothing
Wood
Method GENEPY CHN
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Stone or Glass
Textiles or Clothing
Method of Reflections CHN
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Stone or Glass
Textiles or Clothing
Vegetables
1,0
0,8
0,6
0,4
0,2
0,0
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Sector
Animals
Chemicals
Foodprodu..
Footwear
Fuels
Hides or S..
Machinery ..
Metals
Minerals
Miscellane..
Plastics or ..
Stone or G..
Textiles or ..
Transporta..
Vegetables
Wood

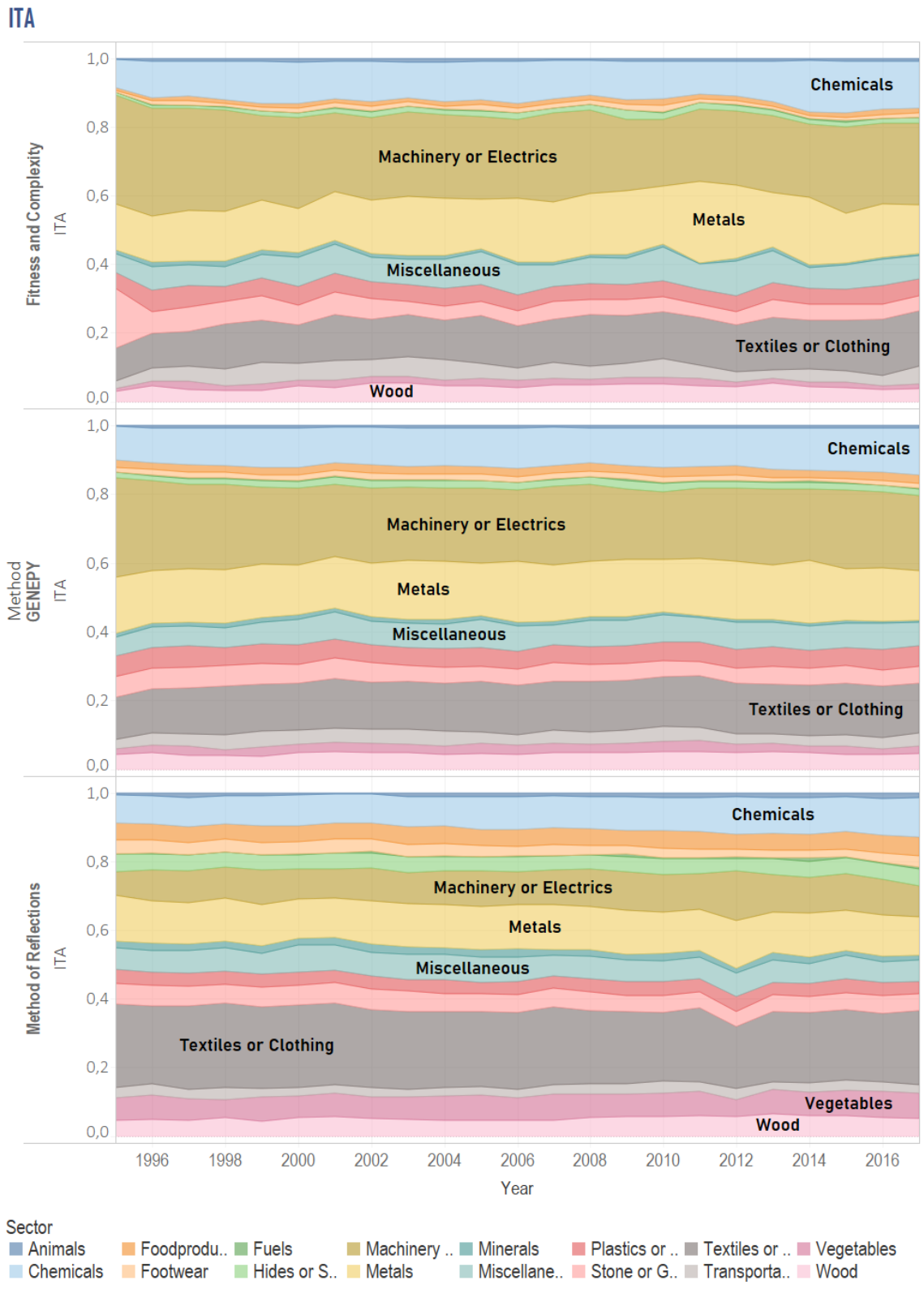
BRA


Fitness and Complexity
BRA
Animals
Chemicals
Machinery or Electrics
Metals
Stone or Glass
Wood
Method GENEPY
BRA
Animals
Chemicals
Foodproducts
Machinery or Electrics
Metals
Stone or Glass
Vegetables
Wood
Method of Reflections
BRA
Animals
Chemicals
Foodproducts
Metals
Minerals
Textiles or Clothing
Vegetables
Wood
1,0
0,8
0,6
0,4
0,2
0,0
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Sector
Animals
Chemicals
Foodprodu..
Footwear
Fuels
Hides or S..
Machinery ..
Metals
Minerals
Miscellane..
Plastics or ..
Stone or G..
Textiles or ..
Transporta..
Vegetables
Wood

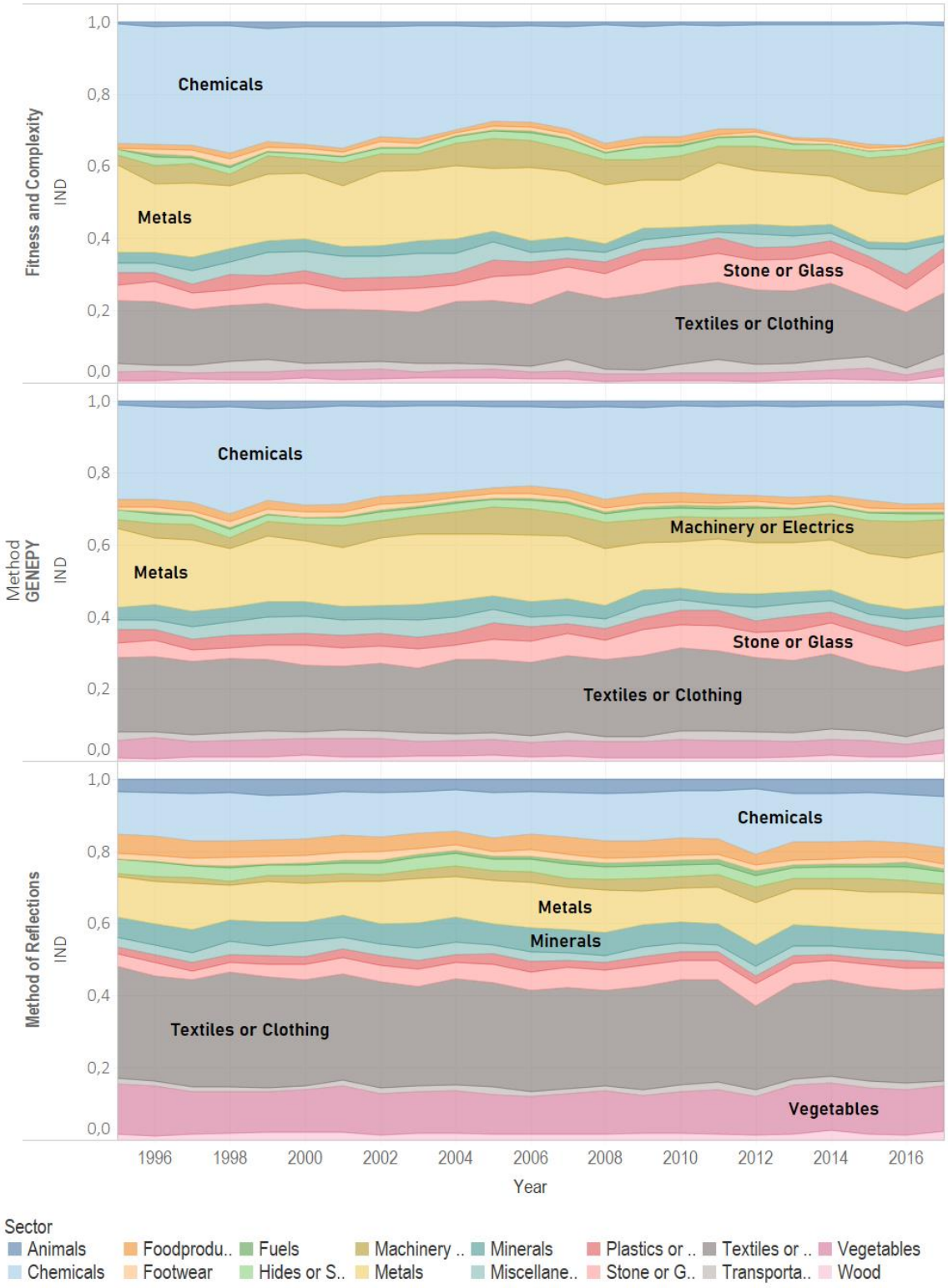

DEU
Fitness and Complexity
DEU
1,0
0,8
0,6
0,4
0,2
0,0
Animals
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Stone or Glass
Method GENEPY
DEU
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Plastics or Rubber
Stone or Glass
Wood
Method of Reflections
DEU
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Textiles or Clothing
Wood
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Sector
Animals
Chemicals
Foodprodu..
Footwear
Fuels
Hides or S..
Machinery ..
Metals
Minerals
Miscellane..
Plastics or ..
Stone or G..
Textiles or ..
Transporta..
Vegetables
Wood

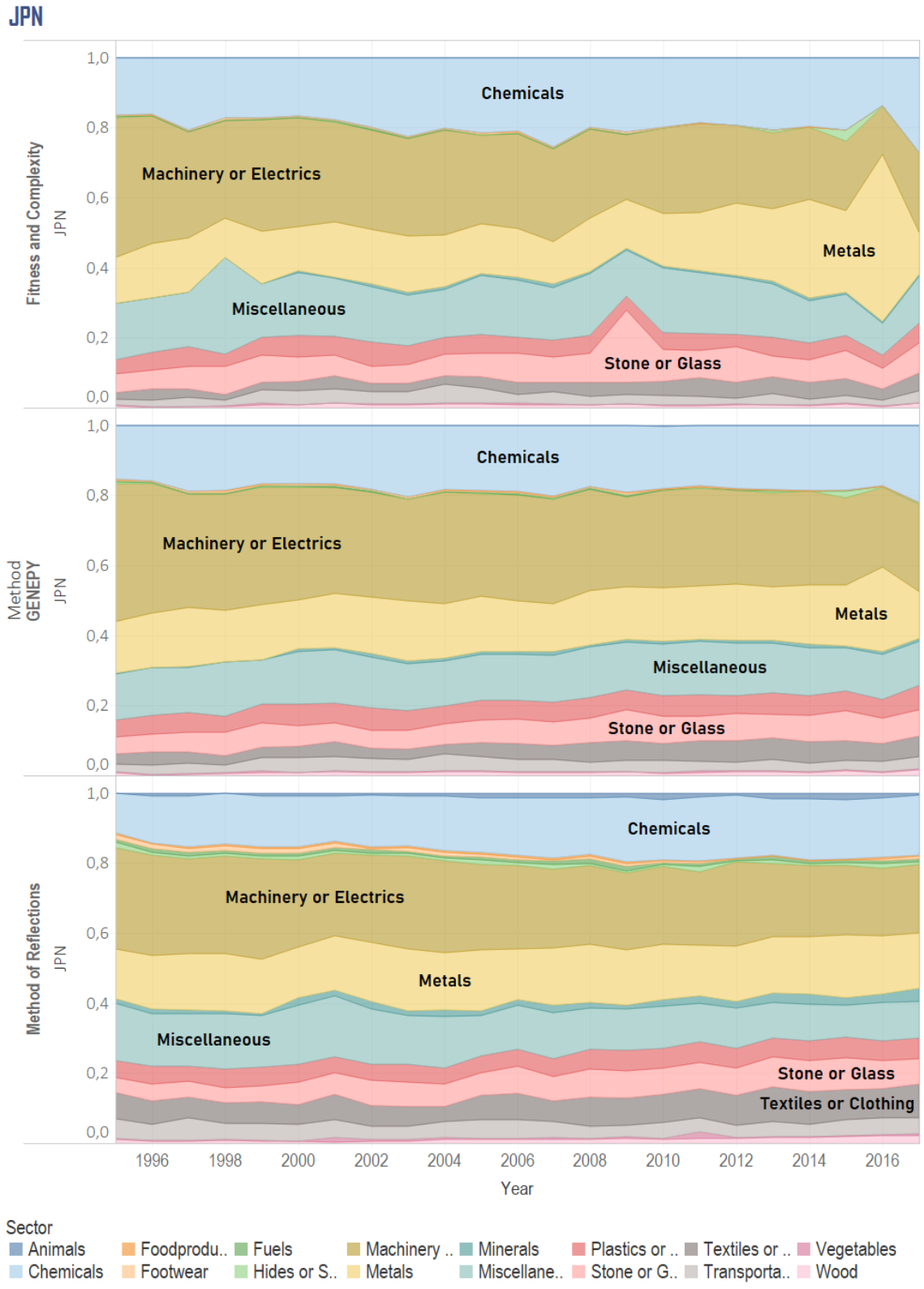
USA


Fitness and Complexity
USA
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Plastics or Rubber
Stone or Glass
Wood
Method GENEPY
USA
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Plastics or Rubber
Stone or Glass
Method of Reflections
USA
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Textiles or Clothing
Vegetables
Wood
1,0
0,8
0,6
0,4
0,2
0,0
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Sector
Animals
Chemicals
Foodprodu..
Footwear
Fuels
Hides or S..
Machinery ..
Metals
Minerals
Miscellane..
Plastics or ..
Stone or G..
Textiles or ..
Transporta..
Vegetables
Wood

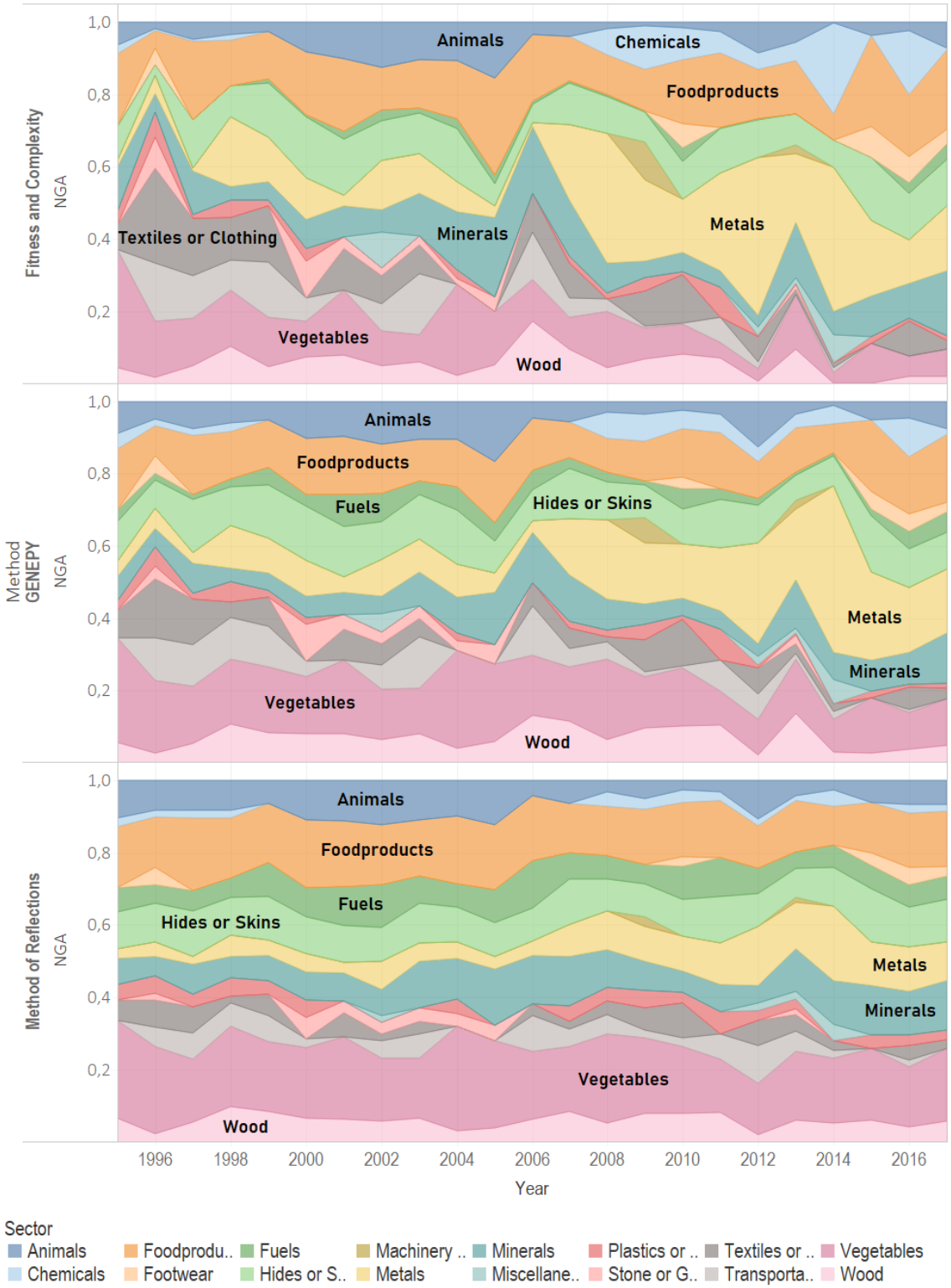

ITA
Fitness and Complexity
ITA
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Textiles or Clothing
Wood
Method
GENEPY
ITA
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Textiles or Clothing
Method of Reflections
ITA
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Textiles or Clothing
Vegetables
Wood
1,0
0,8
0,6
0,4
0,2
0,0
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Sector
Animals
Chemicals
Foodprodu..
Footwear
Fuels
Hides or S..
Machinery ..
Metals
Minerals
Miscellane..
Plastics or ..
Stone or G..
Textiles or ..
Transporta..
Vegetables
Wood

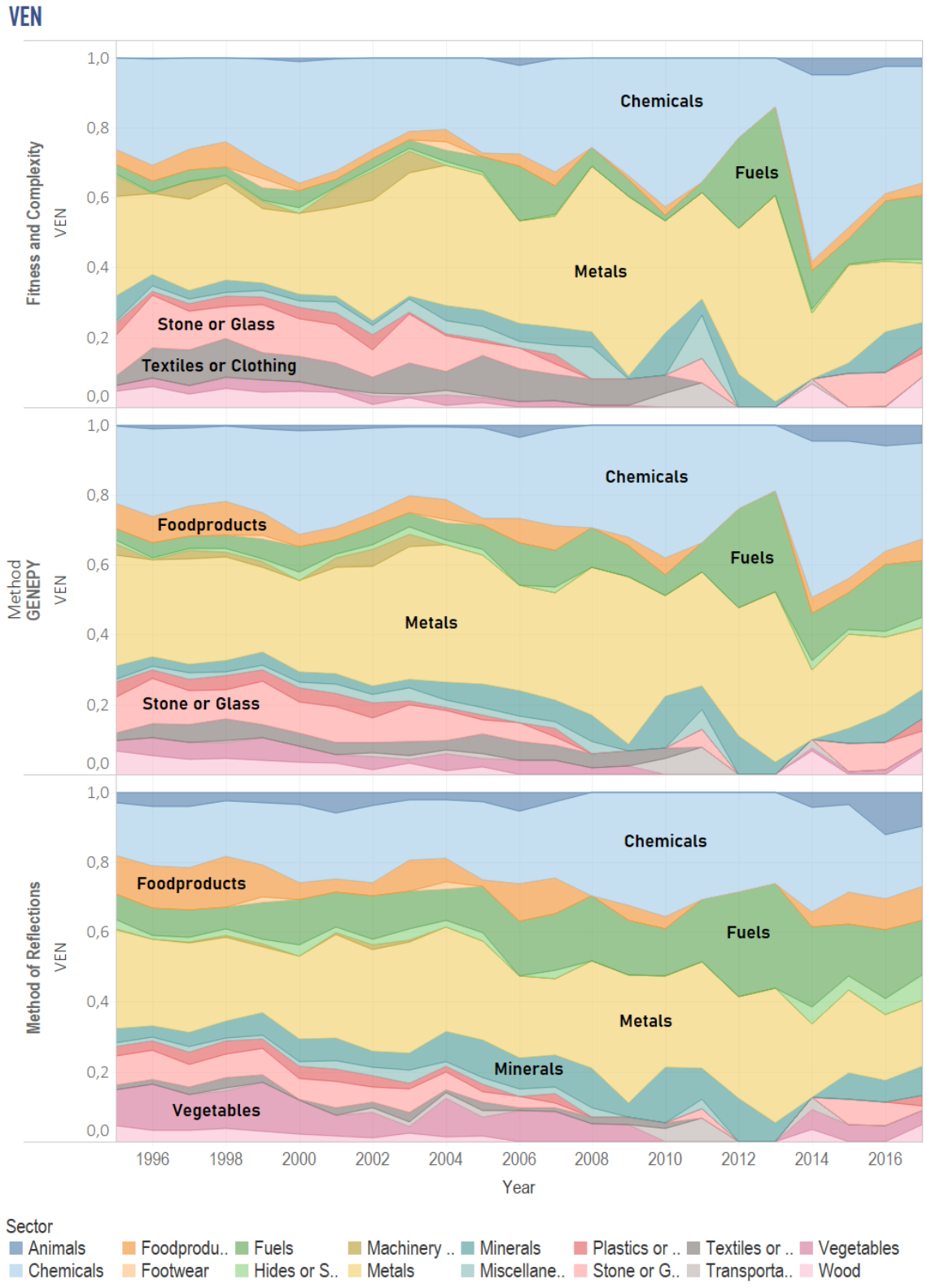
IND


Fitness and Complexity
IND
Chemicals
Metals
Stone or Glass
Textiles or Clothing
Method
GENEPY
IND
Chemicals
Machinery or Electrics
Metals
Stone or Glass
Textiles or Clothing
Method of Reflections
IND
Chemicals
Metals
Minerals
Textiles or Clothing
Vegetables
1,0
0,8
0,6
0,4
0,2
0,0
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Sector
Animals
Chemicals
Foodprodu..
Footwear
Fuels
Hides or S..
Machinery ..
Metals
Minerals
Miscellane..
Plastics or ..
Stone or G..
Textiles or ..
Transporta..
Vegetables
Wood

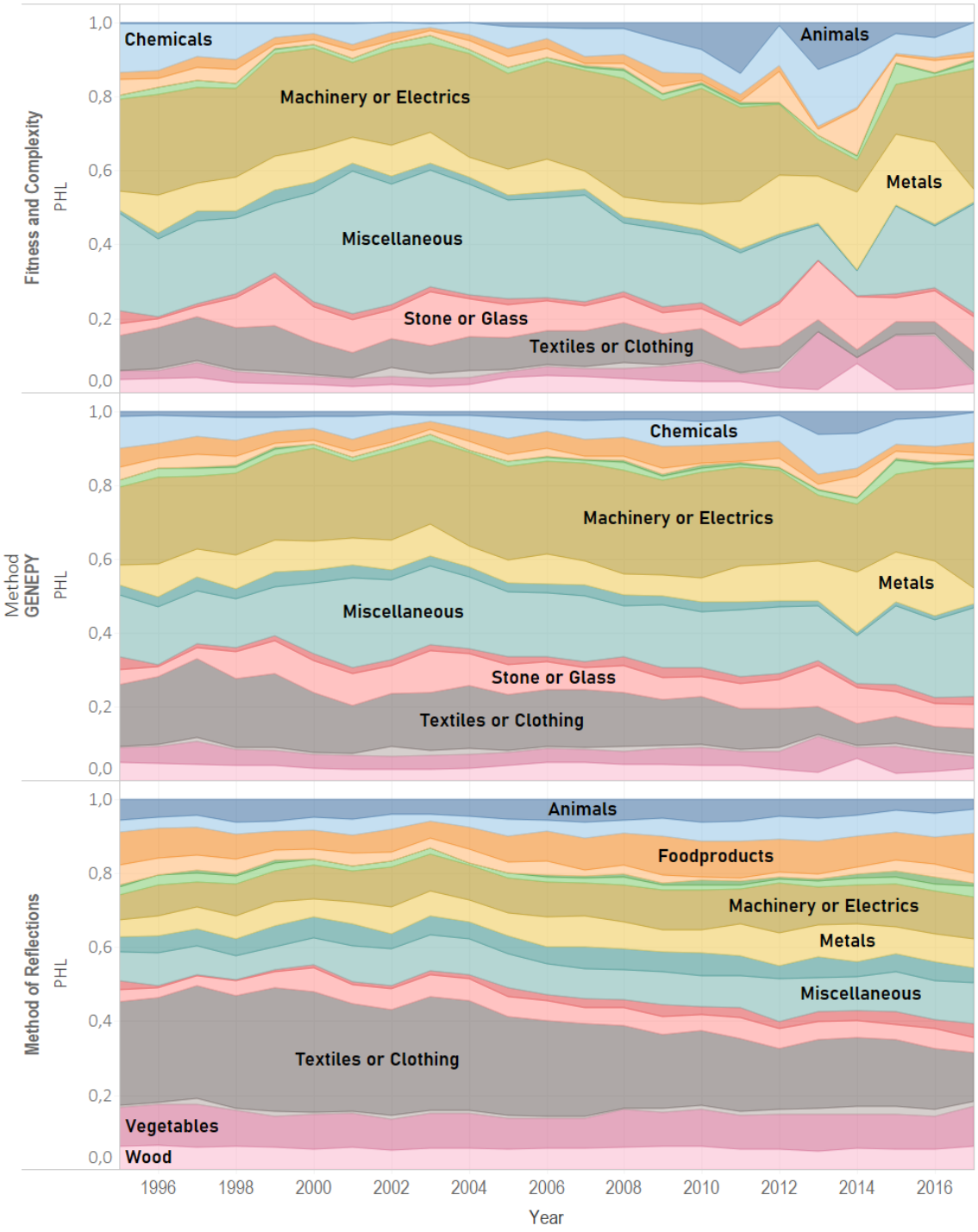

JPN
Fitness and Complexity
JPN
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Stone or Glass
Method GENEPY
JPN
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Stone or Glass
Method of Reflections
JPN
Chemicals
Machinery or Electrics
Metals
Miscellaneous
Stone or Glass
Textiles or Clothing
1,0
0,8
0,6
0,4
0,2
0,0
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Sector
Animals
Chemicals
Foodprodu..
Footwear
Fuels
Hides or S..
Machinery ..
Metals
Minerals
Miscellane..
Plastics or ..
Stone or G..
Textiles or ..
Transporta..
Vegetables
Wood

NGA


Fitness and Complexity
NGA
1,0
0,8
0,6
0,4
0,2
Animals
Chemicals
Foodproducts
Textiles or Clothing
Minerals
Metals
Vegetables
Wood
Method GENEPY
NGA
Animals
Foodproducts
Fuels
Hides or Skins
Metals
Minerals
Vegetables
Wood
Method of Reflections
NGA
Animals
Foodproducts
Fuels
Hides or Skins
Metals
Minerals
Vegetables
Wood
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Sector
Animals
Chemicals
Foodprodu..
Footwear
Fuels
Hides or S..
Machinery ..
Metals
Minerals
Miscellane..
Plastics or ..
Stone or G..
Textiles or ..
Transporta..
Vegetables
Wood

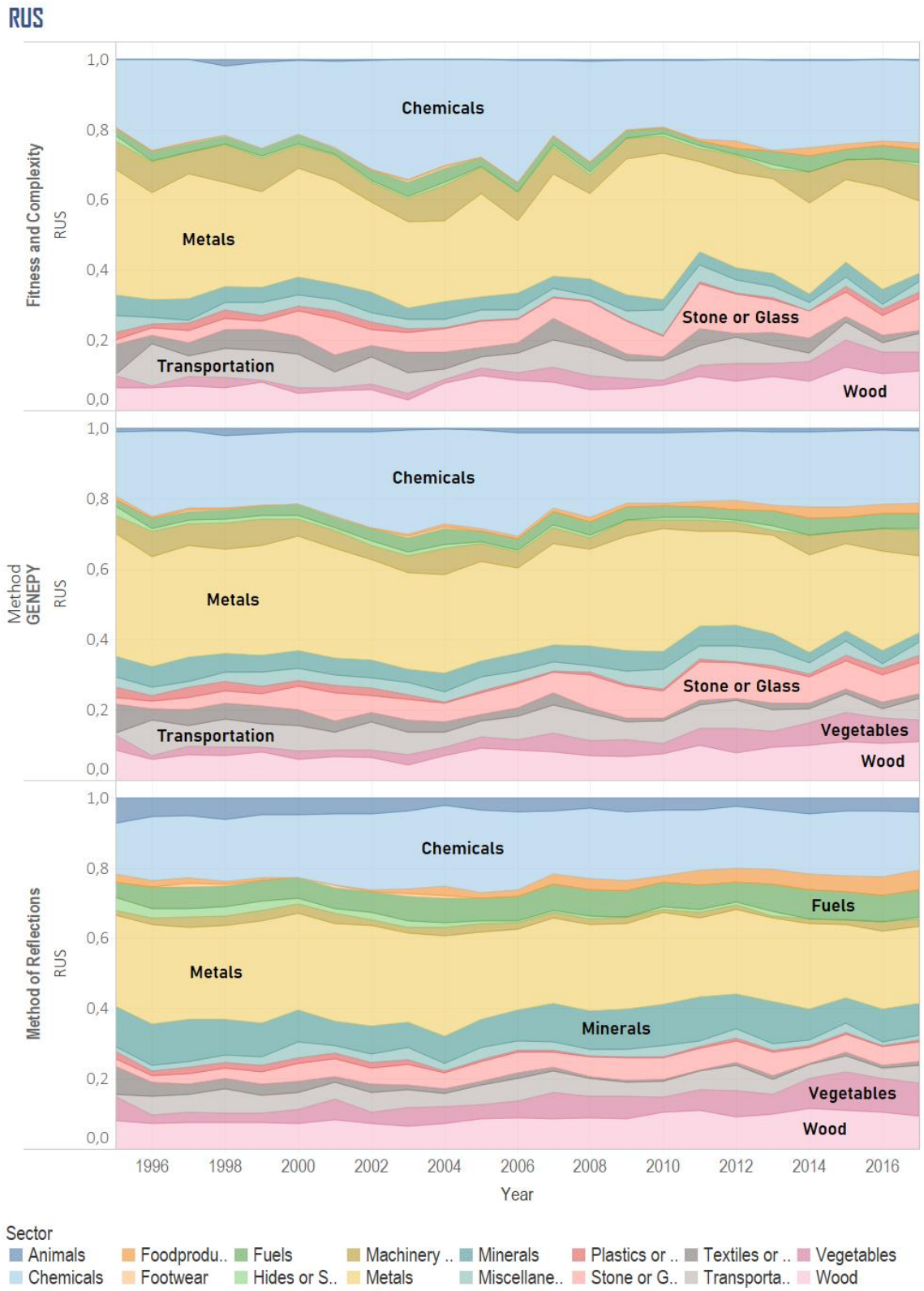

VEN
Fitness and Complexity
VEN
Chemicals
Fuels
Metals
Stone or Glass
Textiles or Clothing
Method GENEPY
VEN
Chemicals
Foodproducts
Fuels
Metals
Stone or Glass
Method of Reflections
VEN
Chemicals
Foodproducts
Fuels
Metals
Minerals
Vegetables
1,0
0,8
0,6
0,4
0,2
0,0
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Sector
Animals
Chemicals
Foodprodu..
Footwear
Fuels
Hides or S..
Machinery ..
Metals
Minerals
Miscellane..
Plastics or ..
Stone or G..
Textiles or ..
Transporta..
Vegetables
Wood

PHL

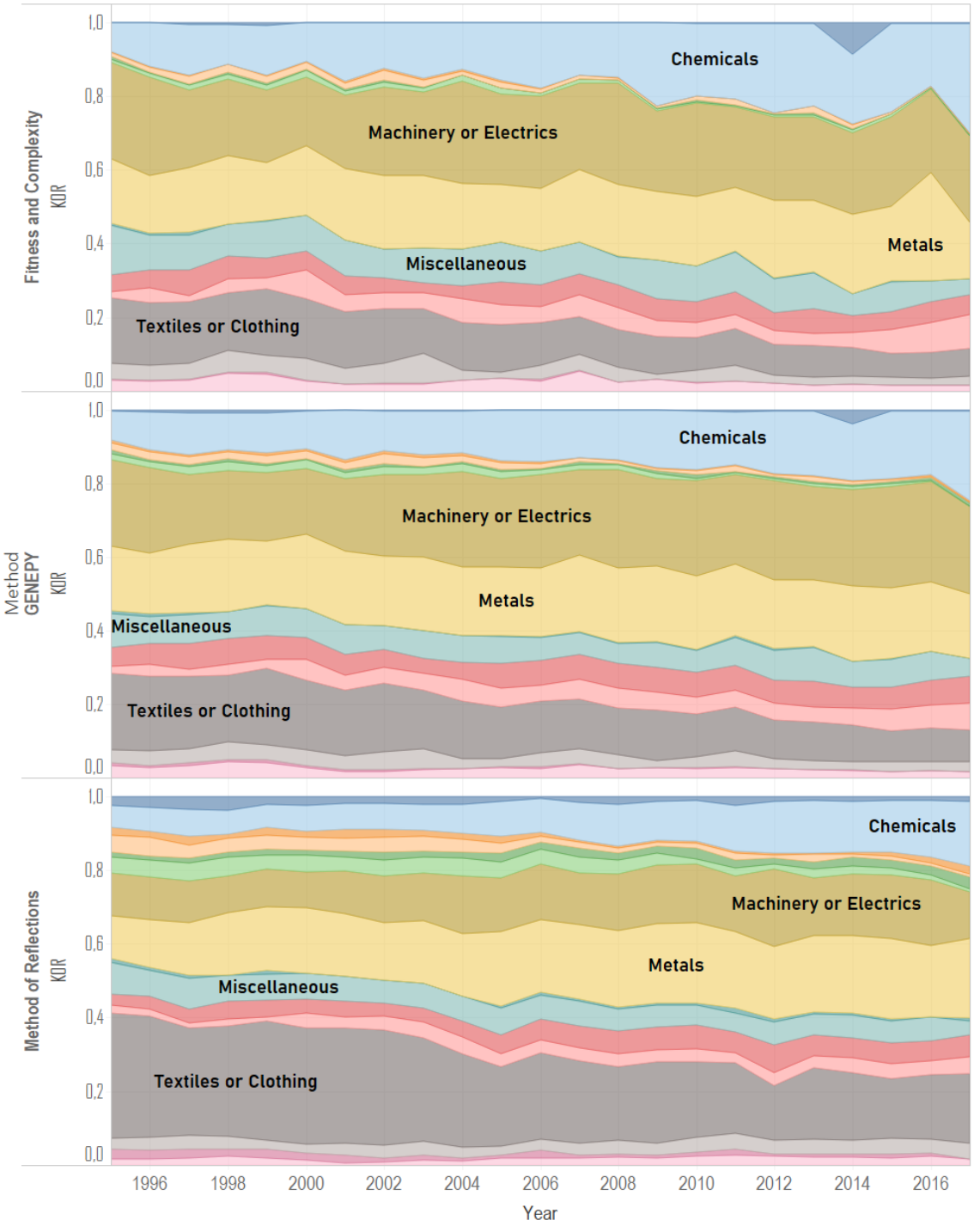


Sector

Animals Chemicals Foodprodu.. Footwear Fuels Hides or S.. Machinery .. Metals Minerals Miscellane.. Plastics or .. Stone or G.. Textiles or .. Transporta.. Vegetables Wood


RUS
Fitness and Complexity
RUS
Chemicals
Metals
Stone or Glass
Transportation
Wood
Method
GENEPY
RUS
Chemicals
Metals
Stone or Glass
Transportation
Vegetables
Wood
Method of Reflections
RUS
Chemicals
Fuels
Metals
Minerals
Vegetables
Wood
1,0
0,8
0,6
0,4
0,2
0,0
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Sector
Animals
Chemicals
Foodprodu..
Footwear
Fuels
Hides or S..
Machinery ..
Metals
Minerals
Miscellane..
Plastics or ..
Stone or G..
Textiles or ..
Transporta..
Vegetables
Wood

KOR



Sector

Animals, Chemicals, Foodprodu.., Footwear, Fuels, Hides or S.., Machinery .., Metals, Minerals, Miscellane.., Plastics or .., Stone or G.., Textiles or .., Transporta.., Vegetables, Wood

Annex A2

China

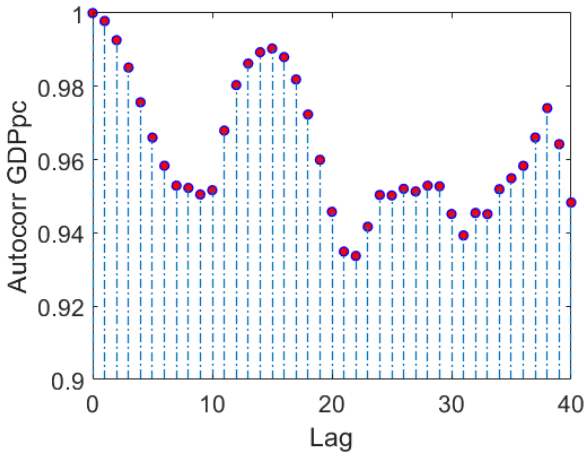


***Figure.A1**: Autocorrelation function of China's GDP time series. GDP time series covers years from 1960 to 2018.*

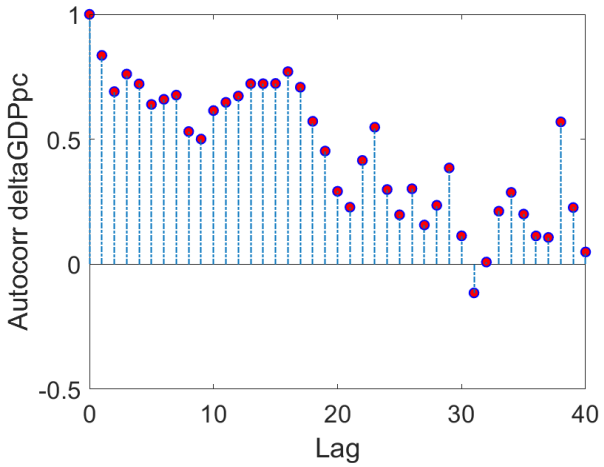


***Figure.A2**: Autocorrelation function of time series of annual increments of GDP (CHN). GDP time series covers years from 1960 to 2018.*

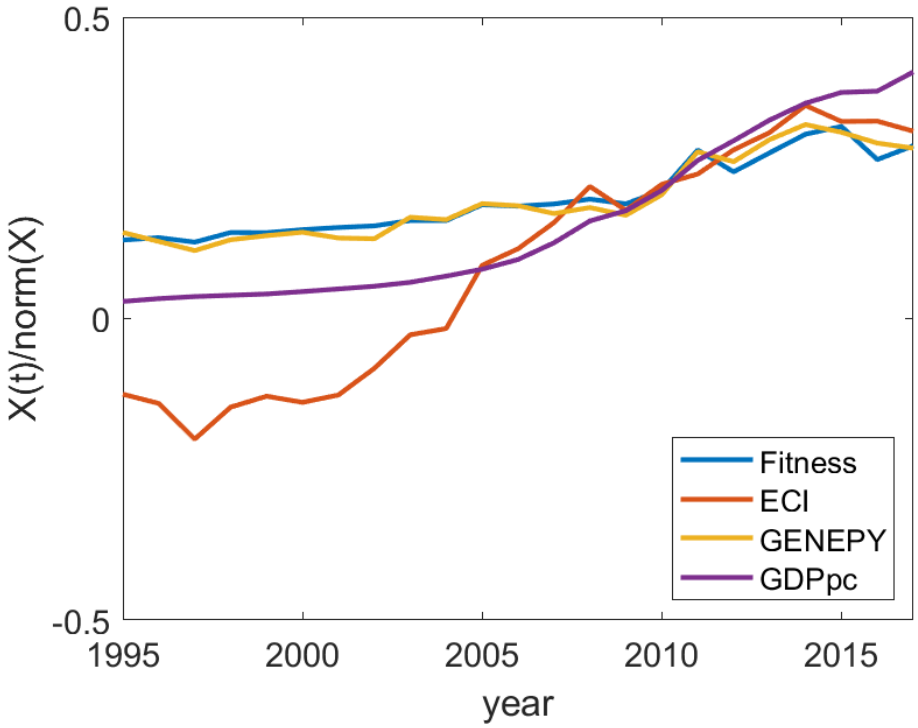


***Figure.A3**: Evolution in time of Economic Indices. Each time series is divided by its Frobenius' norm.*

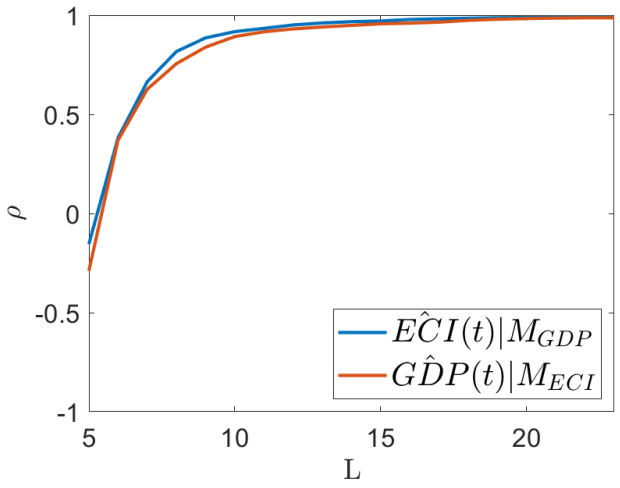


***Figure.A4**: Convergent Cross Mapping between ECI and GDP.*

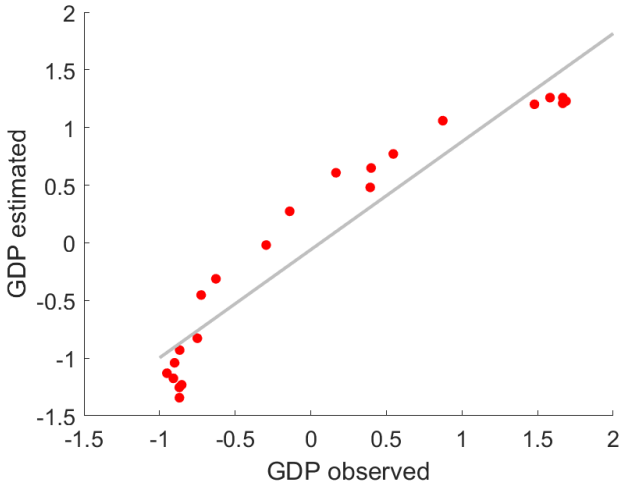


***Figure.A5**: Scatter plot between real time series of GDP per capita and the estimated one by ECI (L=22).*

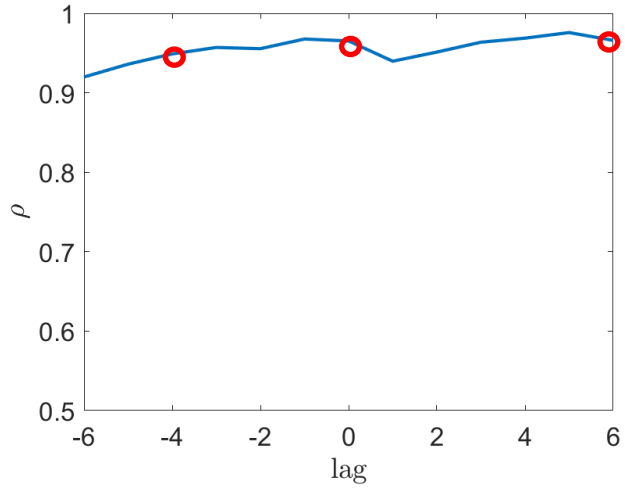


***Figure.A6**: Cross Correlation obtained by reconstruction of GDP at time (t+lag) through ECI at time t.*

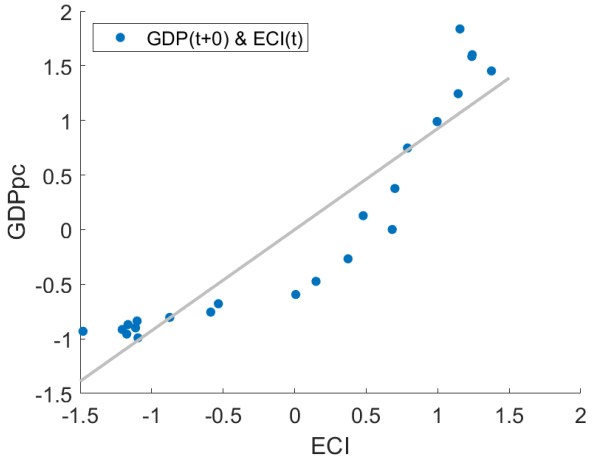


***Figure.A7**: Scatter plot between GDP(t) and ECI(t). Time series are normalized.*

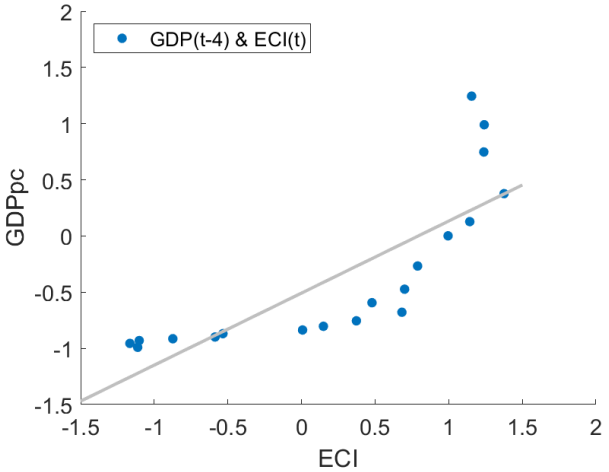


***Figure.A8**: Scatter plot between GDP(t-4) and ECI(t). Time series are normalized.*

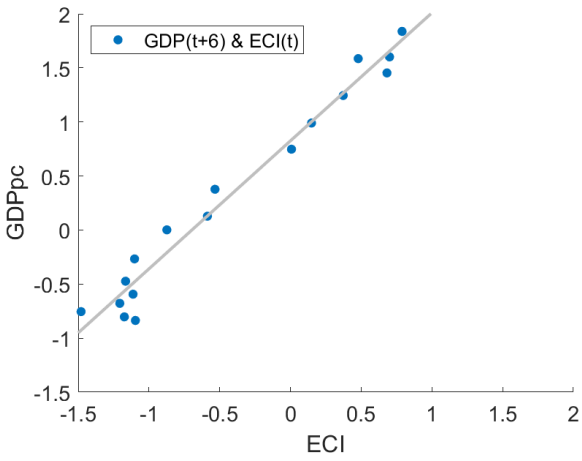


***Figure.A9**: Scatter plot between GDP(t+6) and ECI(t). Time series are normalized.*

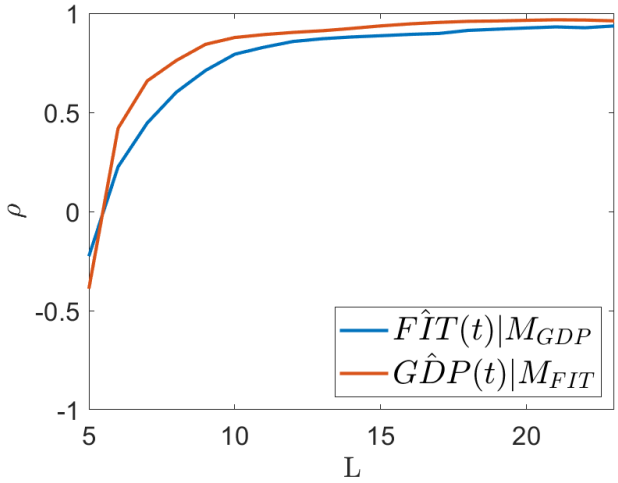


***Figure.A10**: Convergent Cross Mapping between Fitness and GDP.*

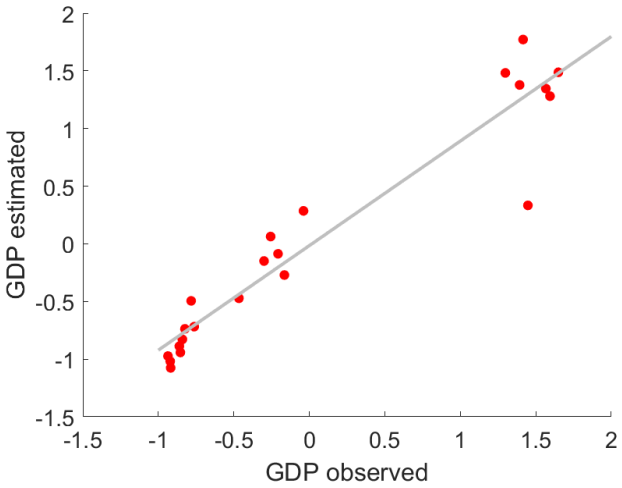


***Figure.A11**: Scatter plot between real time series of GDP per capita and the estimated one by Fitness (L=22).*

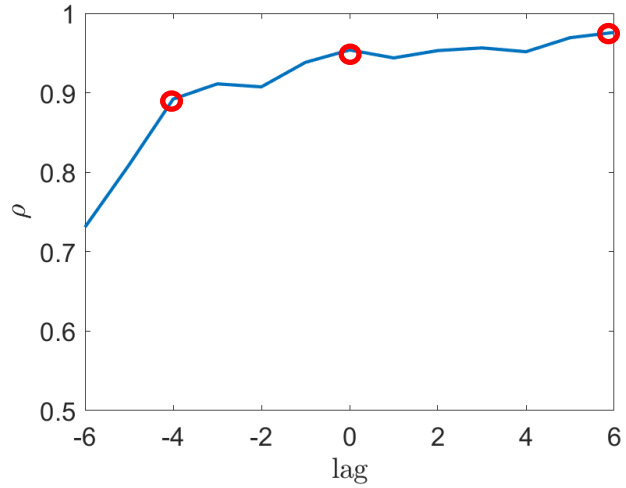


***Figure.A12**: Cross Correlation obtained by reconstruction of GDP at time (t+lag) through Fitness at time t.*

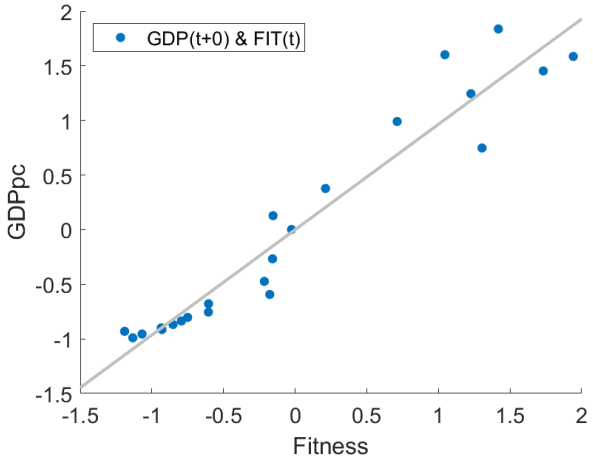


***Figure.A13**: Scatter plot between GDP(t) and Fitness(t). Time series are normalized.*

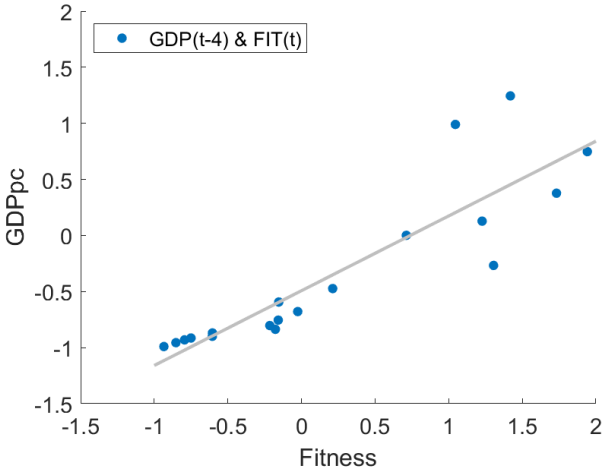


Figure.A14: *Scatter plot between GDP(t-4) and Fitness(t). Time series are normalized.*

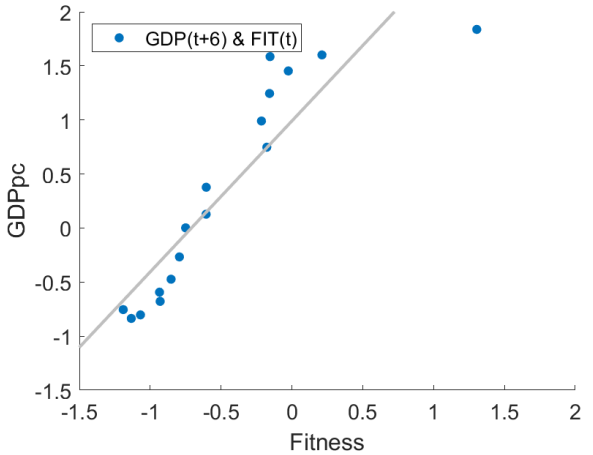


***Figure.A15**: Scatter plot between GDP(t+6) and Fitness(t). Time series are normalized.*

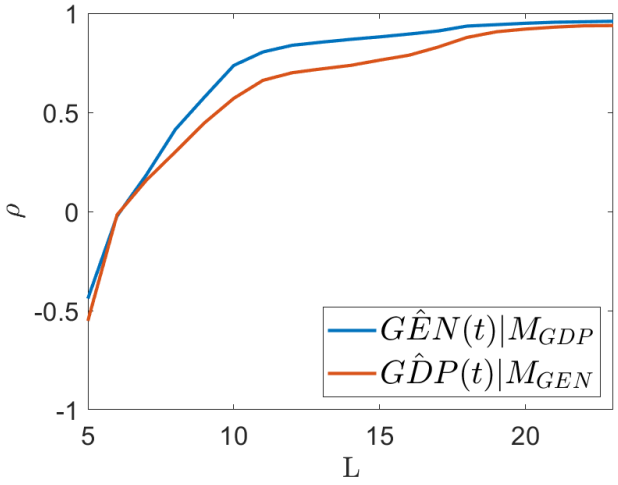


***Figure.A16**: Convergent Cross Mapping between GENEPY and GDP.*

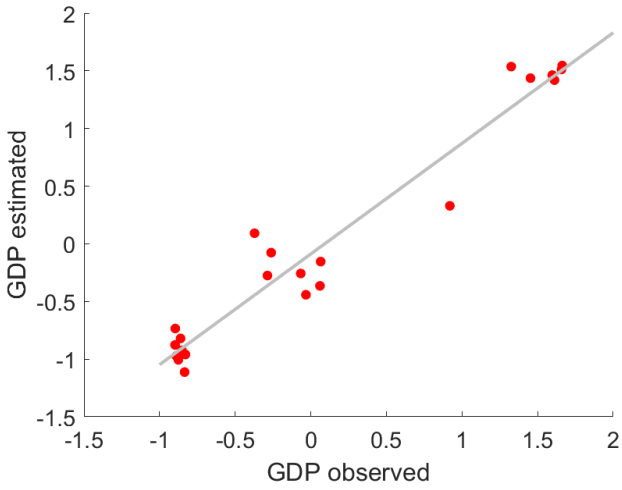


***Figure.A17**: Scatter plot between real time series of GDP per capita and the estimated one by GENEPY (L=22).*

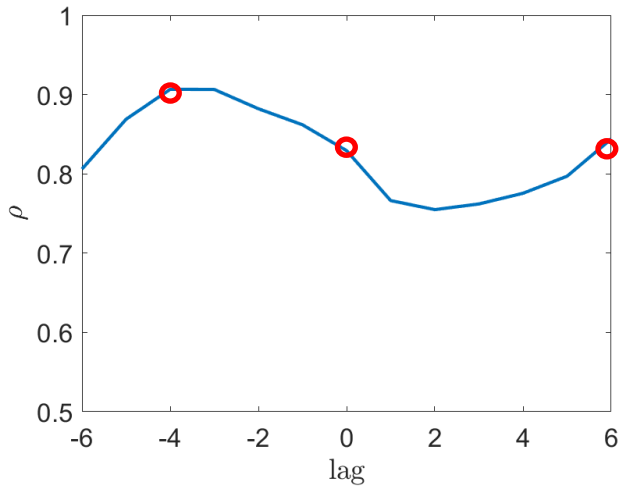


***Figure.A18**: Cross Correlation obtained by reconstruction of GDP at time (t+lag) through GENEPY at time t.*

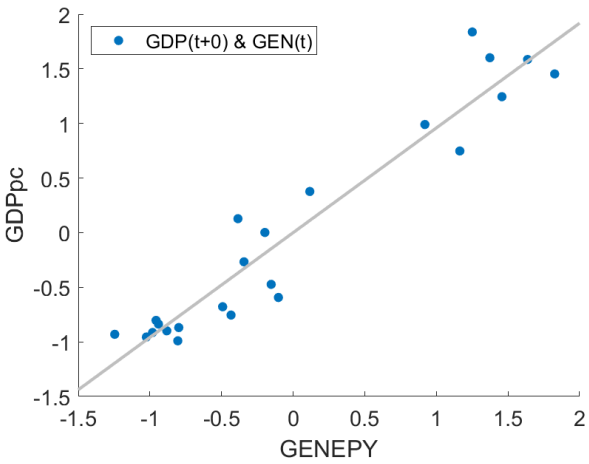


***Figure.A19**: Scatter plot between GDP(t) and GENEPY(t). Time series are normalized.*

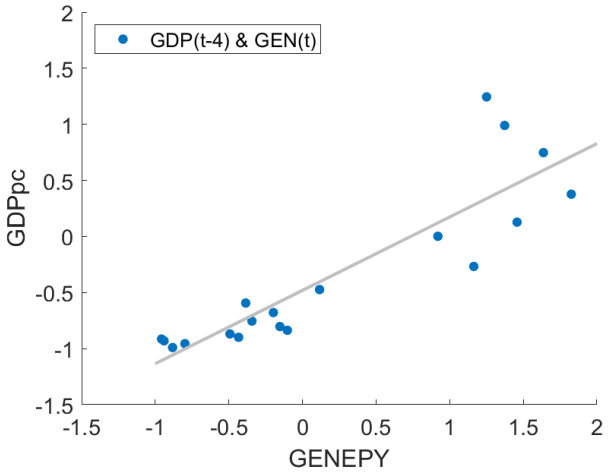


***Figure.A20**: Scatter plot between GDP(t-4) and GENEPY(t). Time series are normalized.*

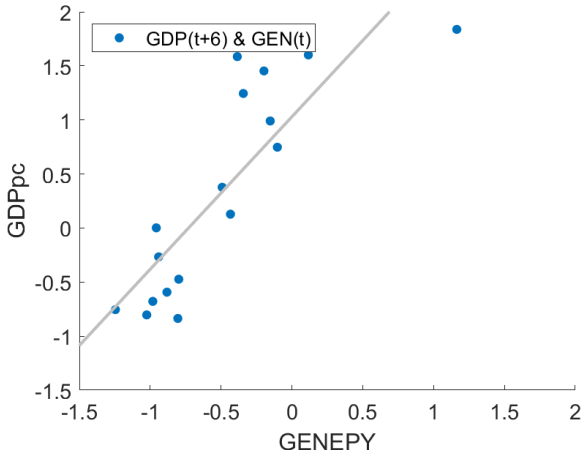


***Figure.A21**: Scatter plot between GDP(t+6) and GENEPY(t). Time series are normalized.*

Italy

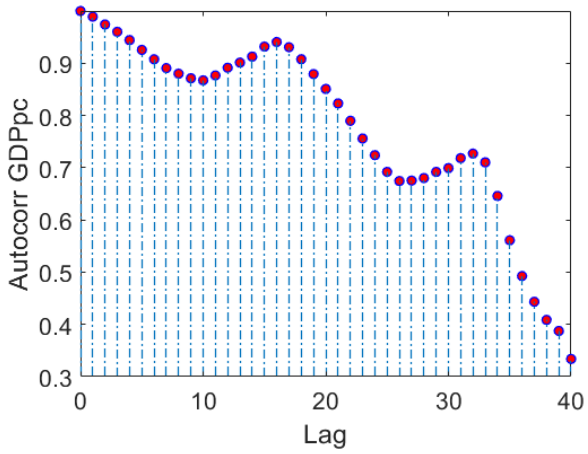


***Figure.A22**: Autocorrelation function of Italy's GDP time series. GDP time series covers years from 1960 to 2018.*

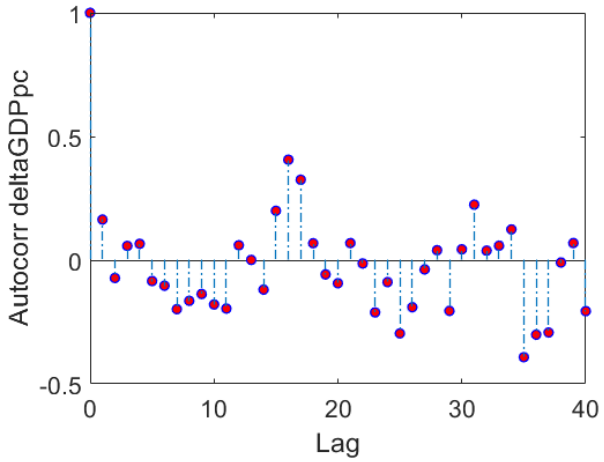


***Figure.A23**: Autocorrelation function of time series of annual increments of GDP (ITA). GDP time series covers years from 1960 to 2018.*

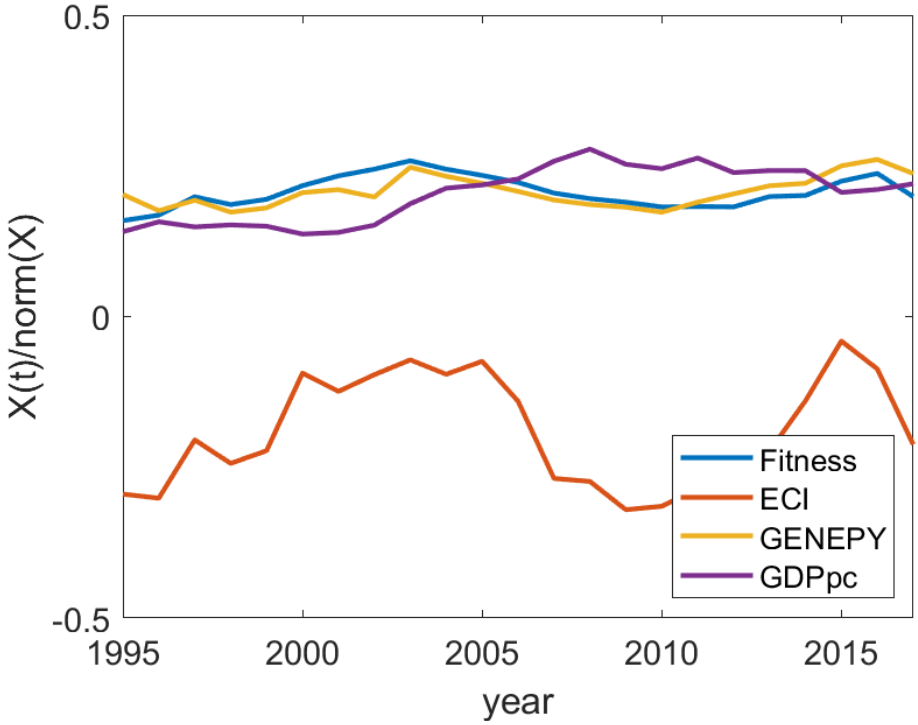


***Figure.A24**: Evolution in time of Economic Indices. Each time series is divided by its Frobenius' norm.*

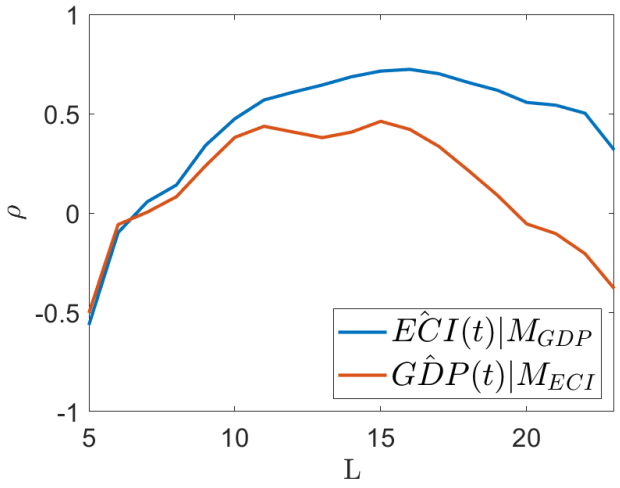


***Figure.A25**: Convergent Cross Mapping between ECI and GDP.*

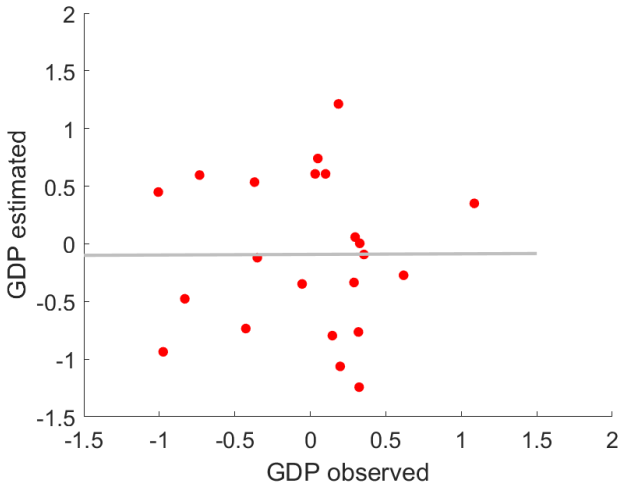


***Figure.A26**: Scatter plot between real time series of GDP per capita and the estimated one by ECI (L=22).*

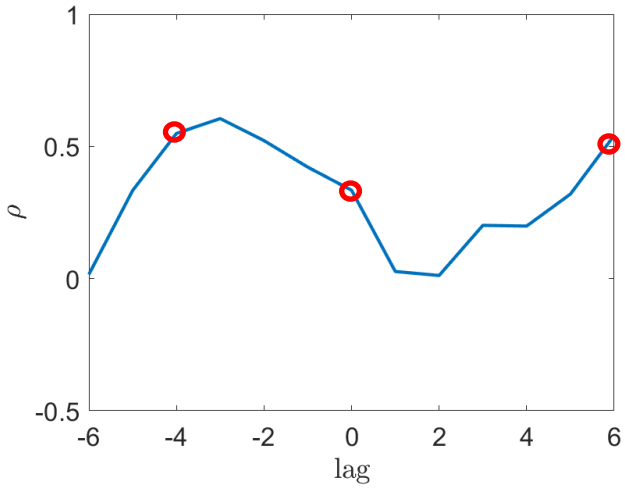


***Figure.A27**: Cross Correlation obtained by reconstruction of GDP at time (t+lag) through ECI at time t.*

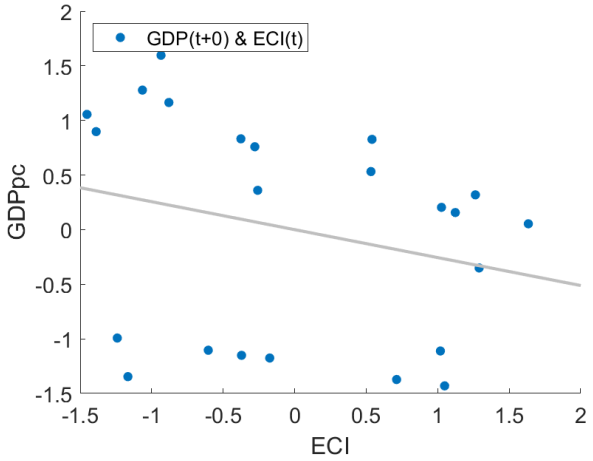


***Figure.A28**: Scatter plot between GDP(t) and ECI(t). Time series are normalized.*

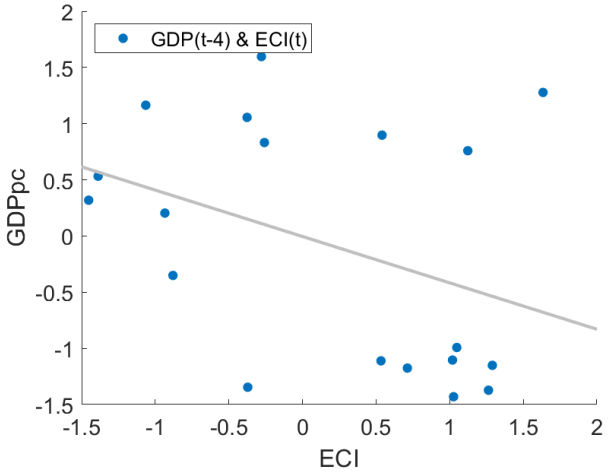


***Figure.A29**: Scatter plot between GDP(t-4) and ECI(t). Time series are normalized.*

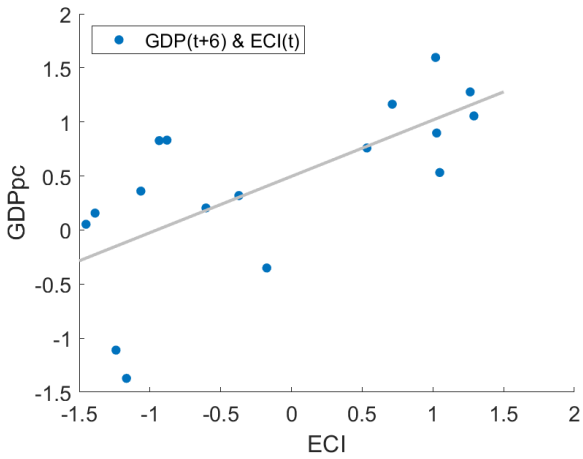


***Figure.A9**: Scatter plot between GDP(t+6) and ECI(t). Time series are normalized.*

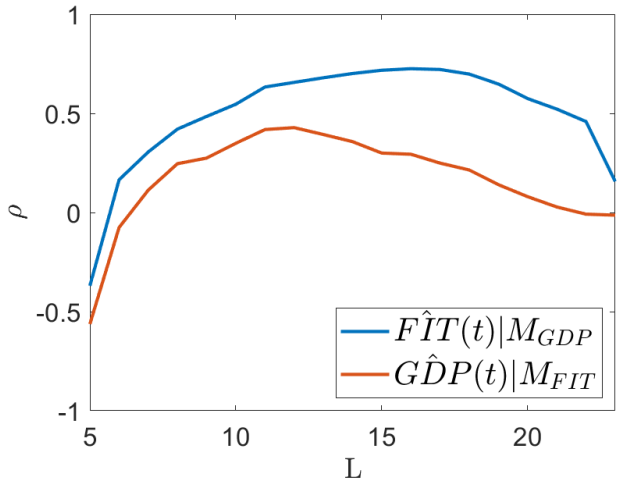


***Figure.A30**: Convergent Cross Mapping between Fitness and GDP.*

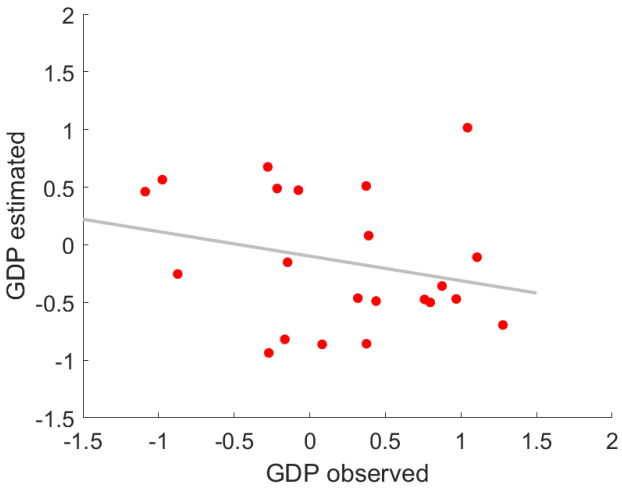


***Figure.A31**: Scatter plot between real time series of GDP per capita and the estimated one by Fitness (L=22).*

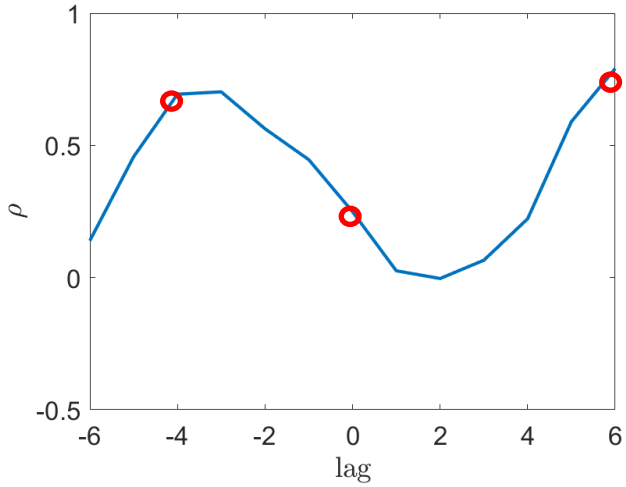


***Figure.A32**: Cross Correlation obtained by reconstruction of GDP at time (t+lag) through Fitness at time t.*

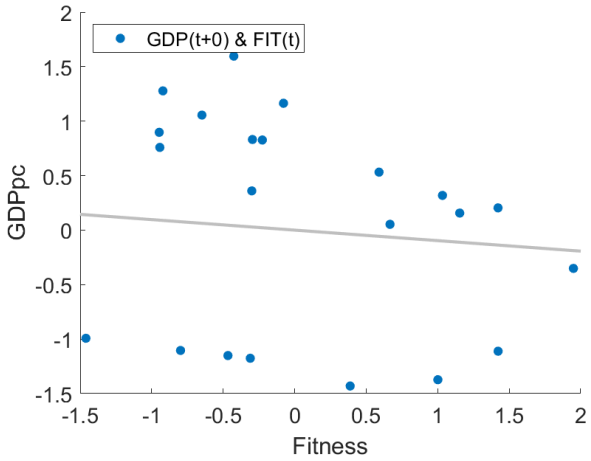


***Figure.A33**: Scatter plot between GDP(t) and Fitness(t). Time series are normalized.*

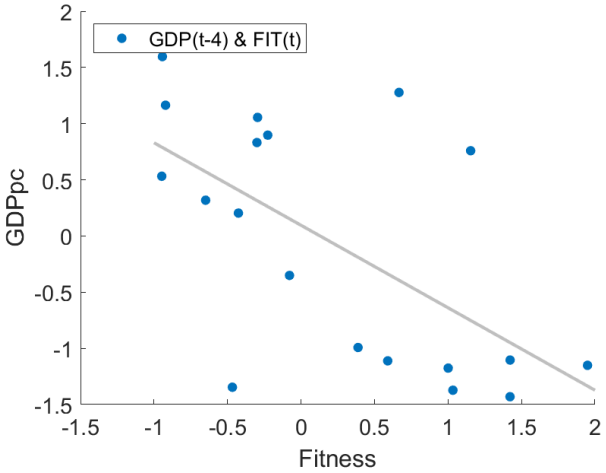


Figure.A34: *Scatter plot between GDP(t-4) and Fitness(t). Time series are normalized.*

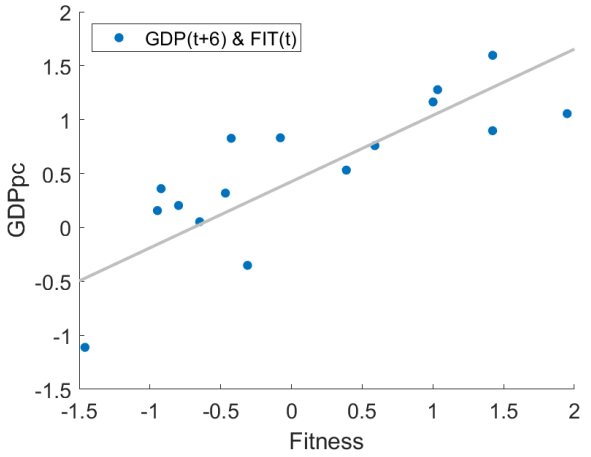


***Figure.A35**: Scatter plot between GDP(t+6) and Fitness(t). Time series are normalized.*

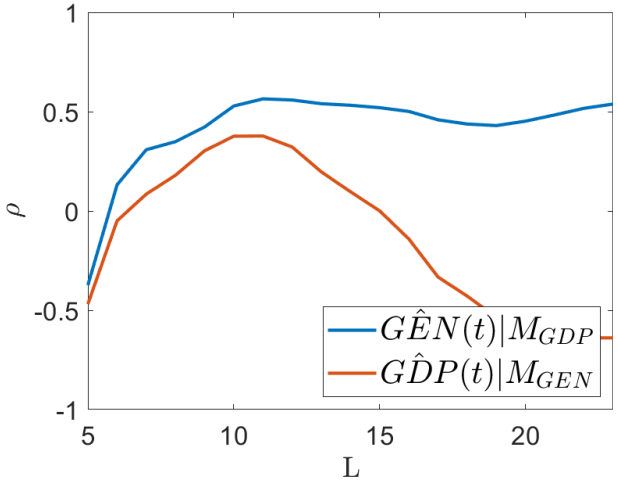


***Figure.A36**: Convergent Cross Mapping between GENEPY and GDP.*

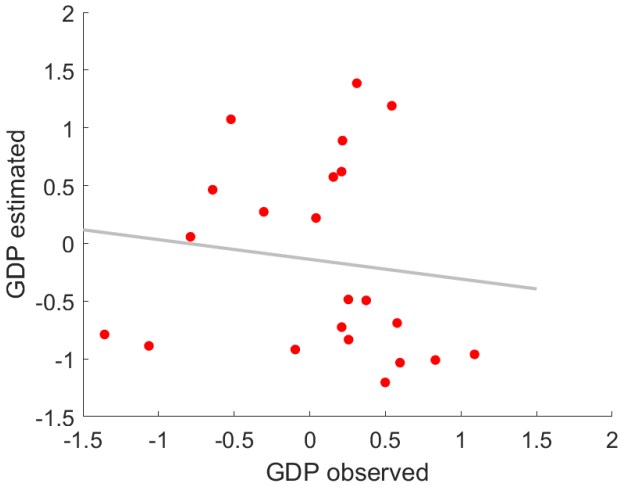


***Figure.A37**: Scatter plot between real time series of GDP per capita and the estimated one by GENEPY (L=22).*

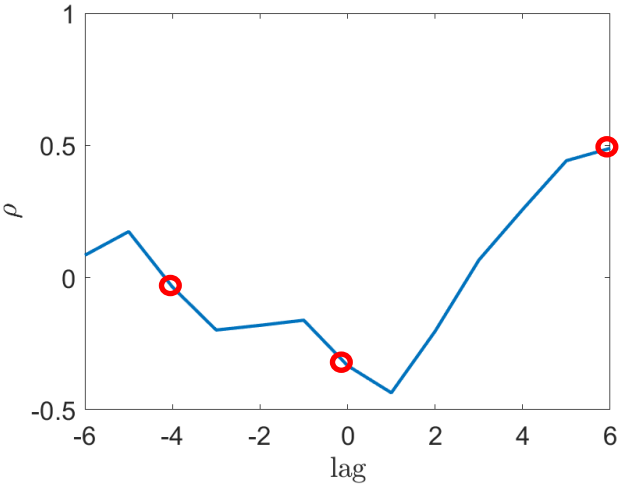


***Figure.A38**: Cross Correlation obtained by reconstruction of GDP at time (t+lag) through GENEPY at time t.*

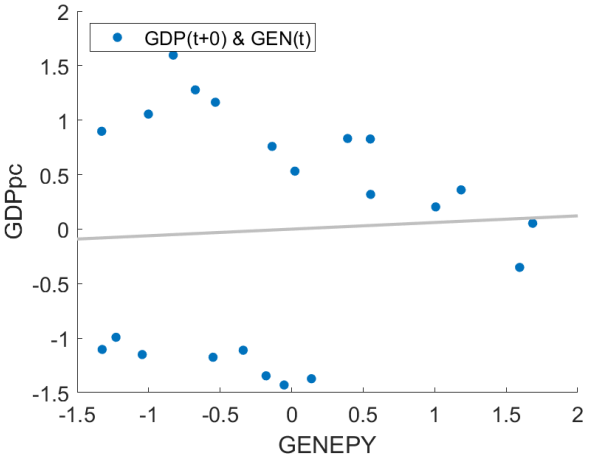


***Figure.A39**: Scatter plot between GDP(t) and GENEPY(t). Time series are normalized.*

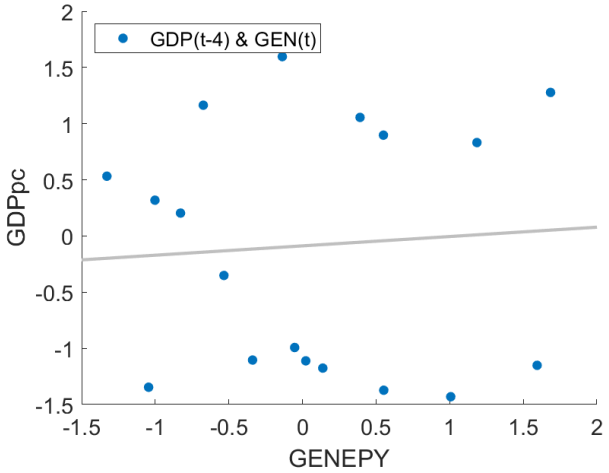


***Figure.A40**: Scatter plot between GDP(t-4) and GENEPY(t). Time series are normalized.*

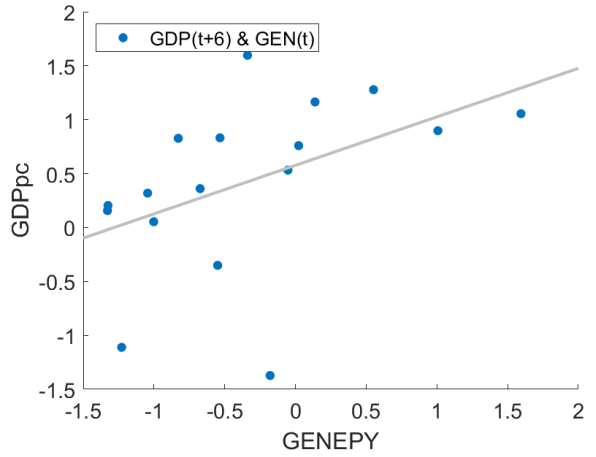


***Figure.A41**: Scatter plot between GDP(t+6) and GENEPY(t). Time series are normalized.*

Singapore

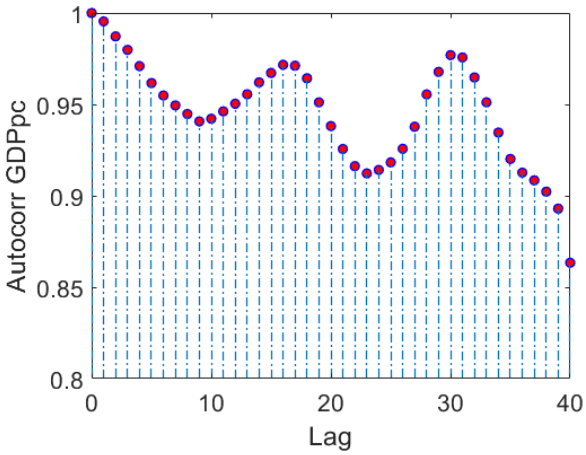


***Figure.A42**: Autocorrelation function of Singapore's GDP time series. GDP time series covers years from 1960 to 2018.*

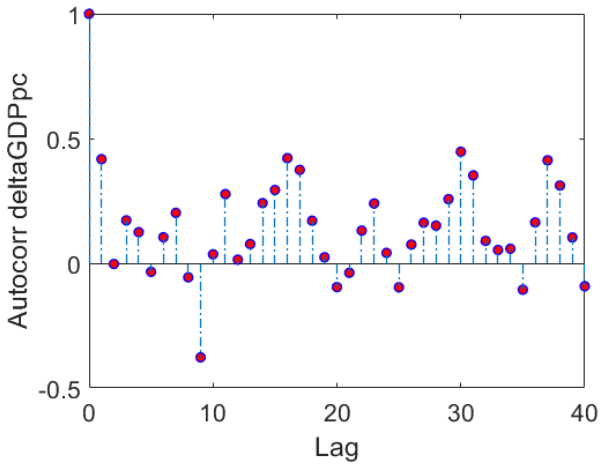


***Figure.A43**: Autocorrelation function of time series of annual increments of GDP (SGP). GDP time series covers years from 1960 to 2018.*

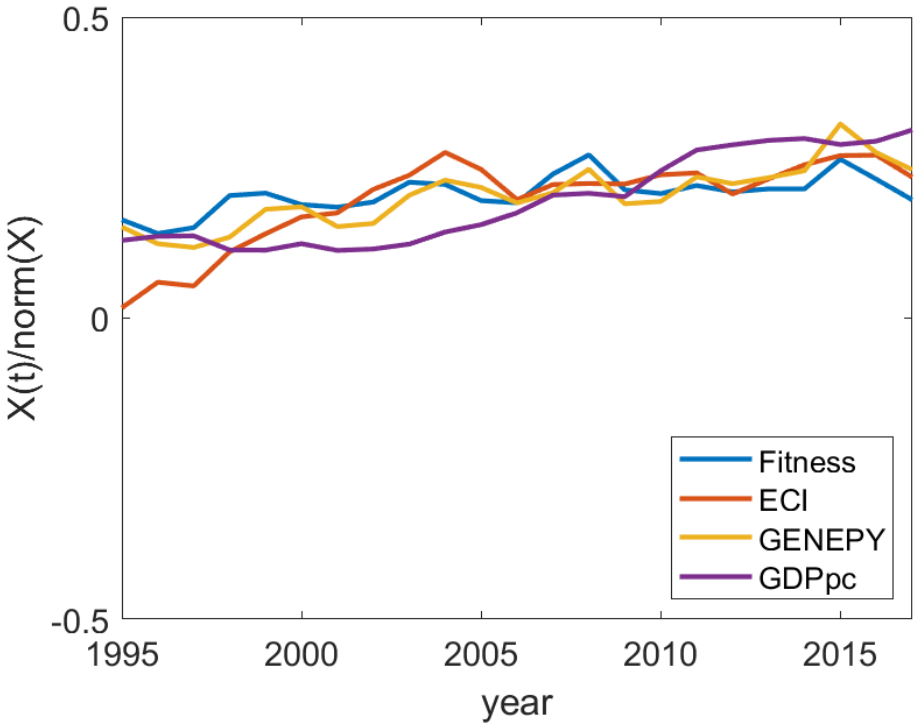


***Figure.A44**: Evolution in time of Economic Indices. Each time series is divided by its Frobenius' norm.*

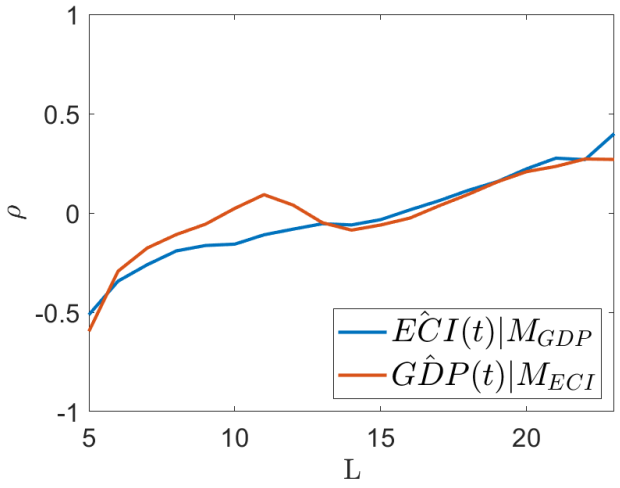


***Figure.A45**: Convergent Cross Mapping between ECI and GDP.*

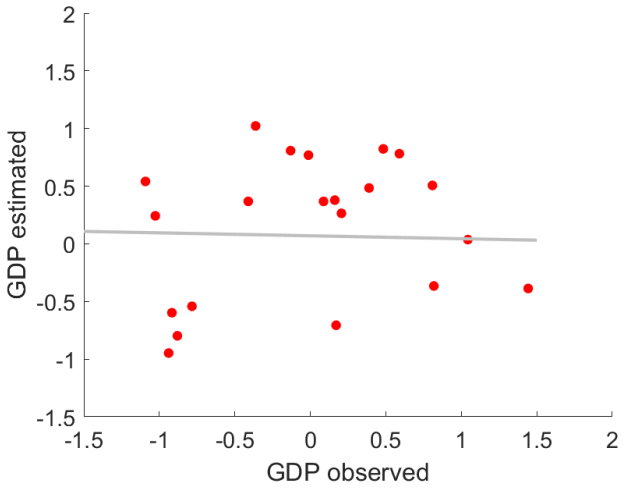


***Figure.A46**: Scatter plot between real time series of GDP per capita and the estimated one by ECI (L=22).*

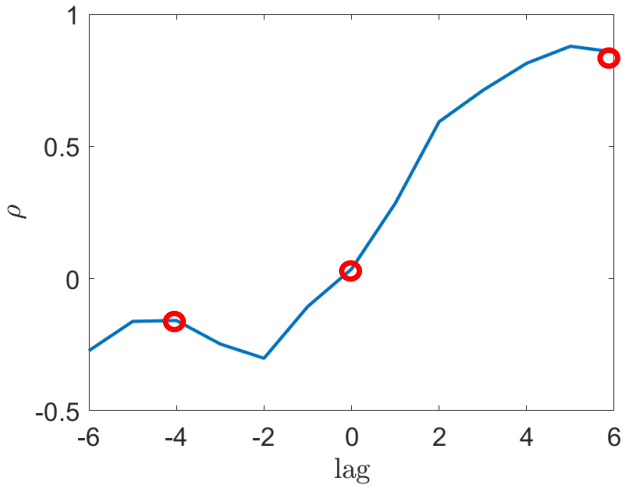


***Figure.A47**: Cross Correlation obtained by reconstruction of GDP at time (t+lag) through ECI at time t.*

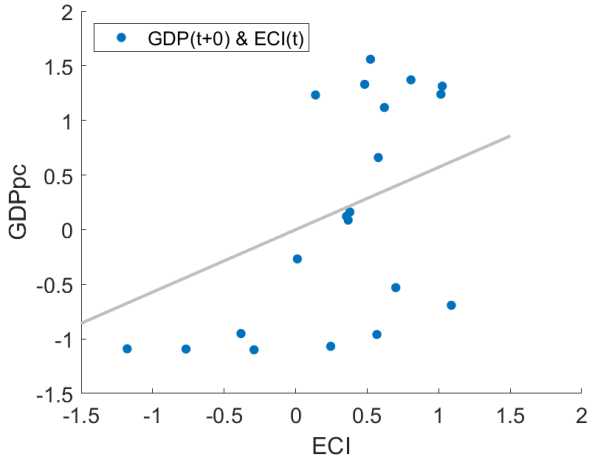


***Figure.A48**: Scatter plot between GDP(t) and ECI(t). Time series are normalized.*

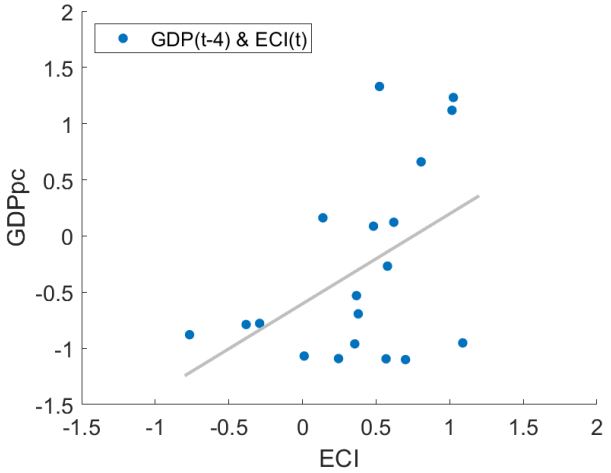


***Figure.A49**: Scatter plot between GDP(t-4) and ECI(t). Time series are normalized.*

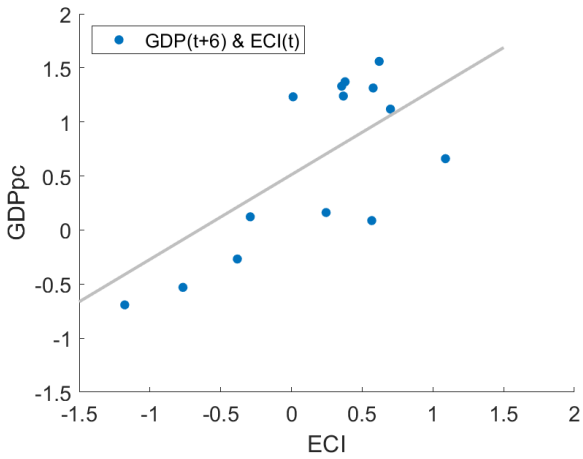


***Figure.A50**: Scatter plot between GDP(t+6) and ECI(t). Time series are normalized.*

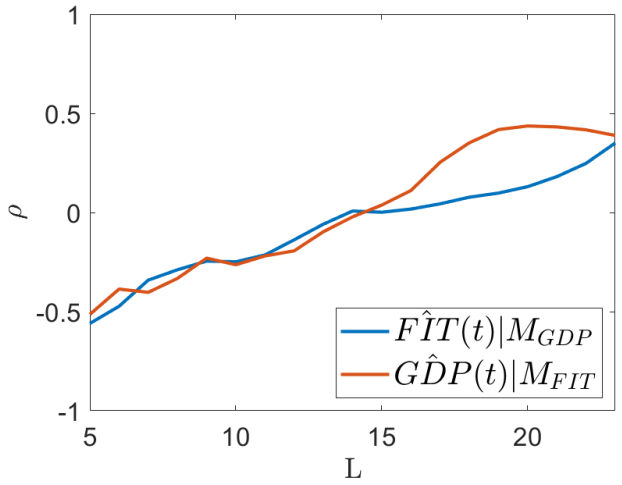


***Figure.A51**: Convergent Cross Mapping between Fitness and GDP.*

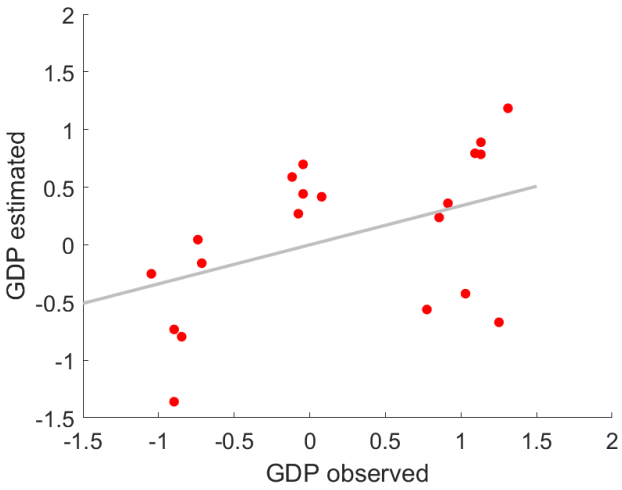


***Figure.A52**: Scatter plot between real time series of GDP per capita and the estimated one by Fitness (L=22).*

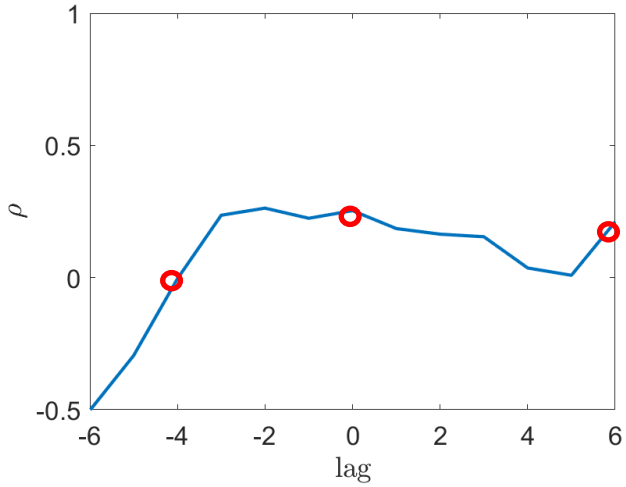


***Figure.A53**: Cross Correlation obtained by reconstruction of GDP at time (t+lag) through Fitness at time t.*

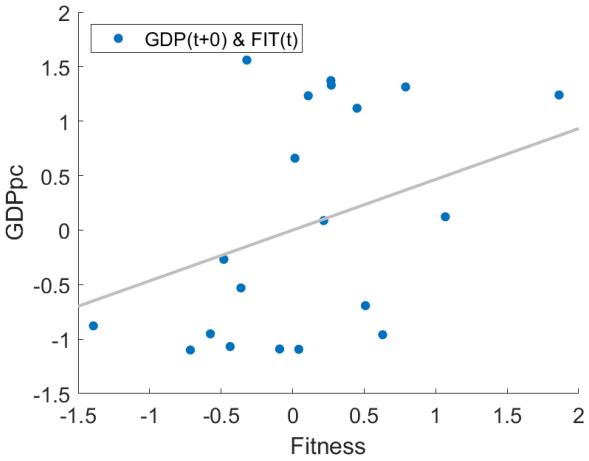


***Figure.A54**: Scatter plot between GDP(t) and Fitness(t). Time series are normalized.*

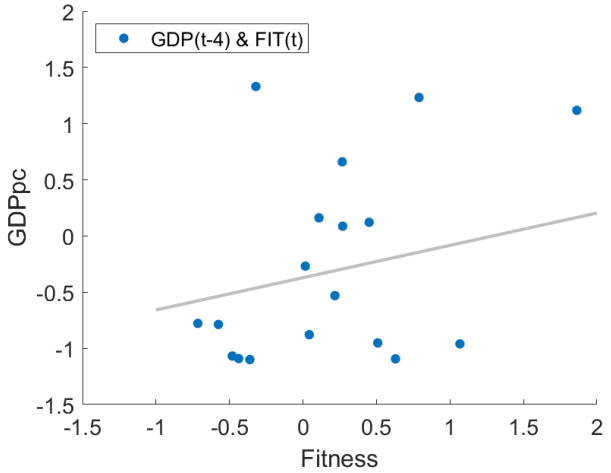


Figure.A55: *Scatter plot between GDP(t-4) and Fitness(t). Time series are normalized.*

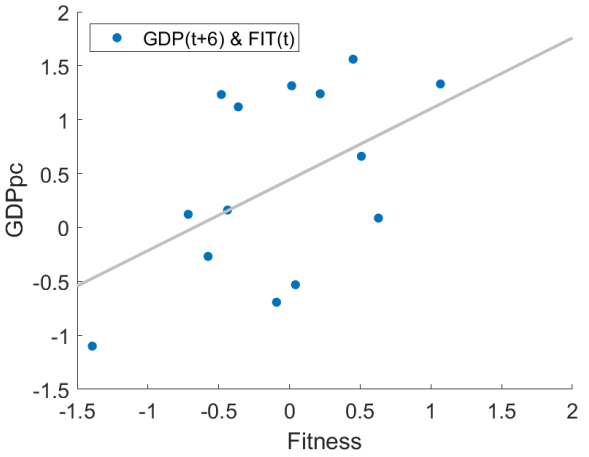


***Figure.A56**: Scatter plot between GDP(t+6) and Fitness(t). Time series are normalized.*

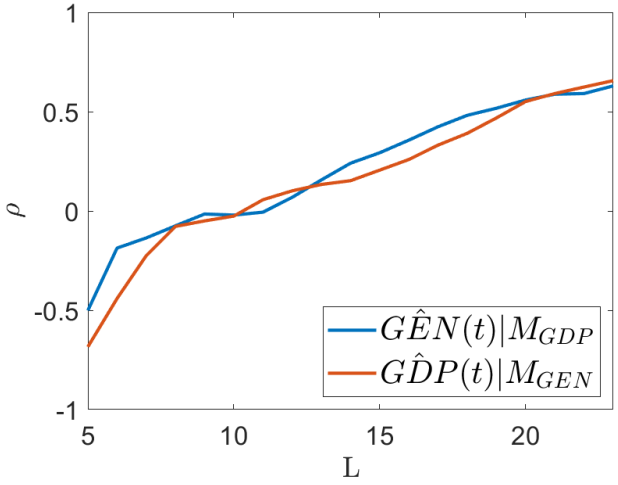


***Figure.A57**: Convergent Cross Mapping between GENEPY and GDP.*

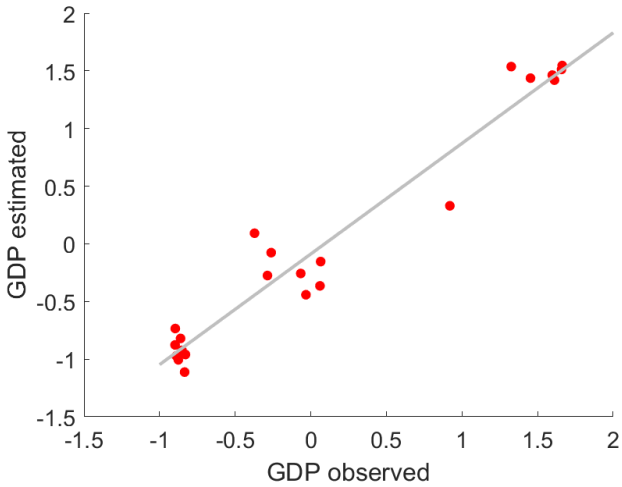


***Figure.A58**: Scatter plot between real time series of GDP per capita and the estimated one by GENEPY (L=22).*

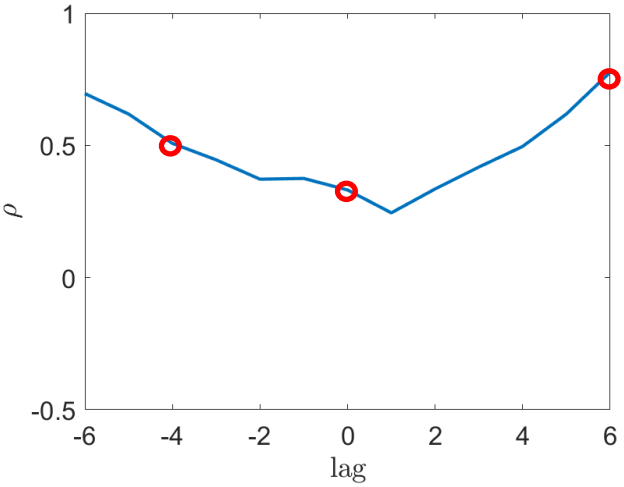


***Figure.A59**: Cross Correlation obtained by reconstruction of GDP at time (t+lag) through GENEPY at time t.*

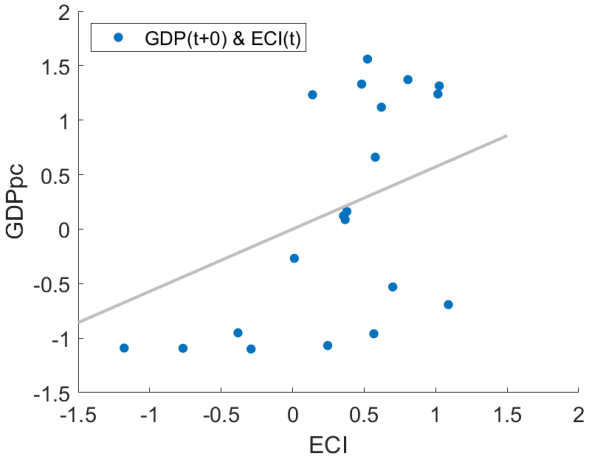


***Figure.A60**: Scatter plot between GDP(t) and GENEPY(t). Time series are normalized.*

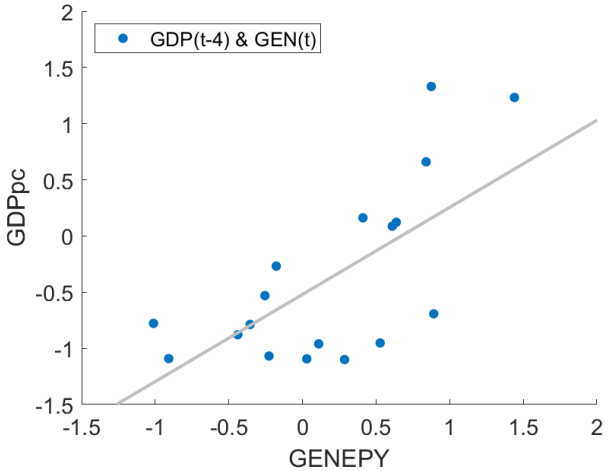


***Figure.A61**: Scatter plot between GDP(t-4) and GENEPY(t). Time series are normalized.*

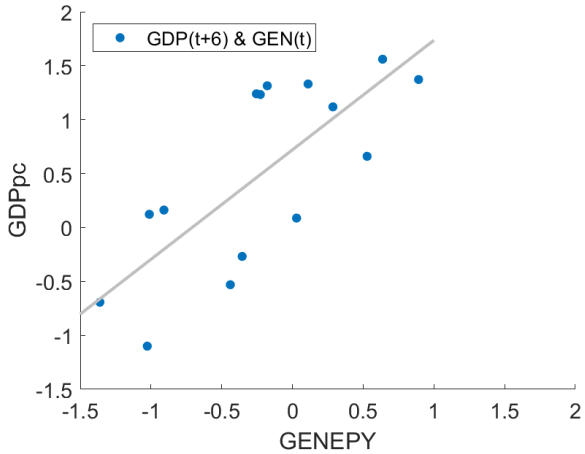


***Figure.A62**: Scatter plot between GDP(t+6) and GENEPY(t). Time series are normalized.*