UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics

Final Dissertation

Economic efficiency and wealth prospects in the global network of countries

Thesis supervisor
Prof. Fulvio Baldovin
Thesis co-supervisors
Prof. Attilio L. Stella
Dr. Gianluca Teza

Candidate
Andrea Andreini

Academic Year 2021/2022

Contents

1 Introduction

Throughout the last decades, the rapid evolution of the branch of research dedicated to the study of complex systems has led to a significant integration of physics into disparate scientific fields, including microbiology, neurology, ecology, finance and many others. As a matter of fact, concepts and methods which are distinctive to the physics approach have indeed proven their ability to supply a deeper understanding of a heterogeneous set of systems, specifically reached through an effecting modeling of their internal structure and of the interactions among their components. A notable example of this circumstance is provided by the economic complexity scenario, where in recent years several peculiar physical models have been proposed for the description of the complex web of interactions between the countries and the global markets and have led to a substantial development of the econophysics branch of research.

The main purpose of this work lies precisely in this context and consists in the characterization of the quantitative nexus which links the economic growth of a country to the evolution in time of its trades and to the overall diversification of its commercial activity. Specifically, these last two aspects will be investigated through the implementation of, respectively, an innovative minimal dynamic model consisting in a set of coupled stochastic differential equations (SDE) and a set of novel complexity measures derived from a refinement of the celebrated Shannon's entropy function; ultimately, the unification of these two methods of analysis will supply a measure of the relative economic efficiency of the countries in the international scenario and will open the possibility to realize meaningful assessments of their wealth prospects.

In the context of economic complexity, network structures have shown a marked ability to provide a straightforward and intuitive depiction of the intertwined nature that often characterizes the systems being analyzed. The employment of this kind of structures has characterized this branch of research since its dawn: among the most relevant examples introduced over time are the *World Trade Web* by Angeles Serrano and Boguná [1], a network of countries defined in order to represent the international flux of money, and the *Product Space* by Hausmann and Klinger [2], a network designed for the characterization of the similarity between the trades of different products. The topology of the feasible networks is quite variegated and reflects the great variety of problems affordable through this kind of approach. In particular, the topological structure which subtends the whole analysis presented in this work is a double-layered bipartite network. Specifically, the nodes that constitute the two layers represent, respectively, the countries of the World and the products traded on the global market; moreover, as suggested by the intuition, a link connecting two nodes lying on different layers reflects the presence of a given product inside either the export or the import basket of the selected country and its strength is determined by the monetary amount of the trade. This characterization of the links, extensively employed in this work, allows to exploit all the information provided by a trading database; however, threshold criteria for the selection of the most relevant products (in a certain sense) contained in a given basket (like the frequently employed *Revealed Comparative Advantage*) can also be applied [3], depending on the purpose of the analysis. Starting

from this theoretical framework, the product nodes connected to a fixed country (or to a fixed aggregate of countries) can be isolated in order to create a sub-network of the aforementioned Product Space, with the purpose of describing the internal properties of the respective baskets. In this specific framework, the technical characterization of the links between the product nodes plays a role of primary importance. For example, the original analysis proposed by Hausmann and Klinger in [2] quantifies the strength of such links through the implementation of a *proximity matrix*, whose entries are defined as probabilities of the generation of novel trades conditioned on the existence of other similar ones. More precisely, the proximity matrix presents a symmetrical structure (fixing two trades, only the minimum between the aforementioned probabilities is considered) and the similarity between the trades is defined only from factors related to the respective productive infrastructures: for these reasons, a major limitation to the employment of the proximity concept comes from the impossibility to properly describe the correlation between trades that are distant from the productive point of view or that are related by a significant unidirectional causal relation.

The dynamic model employed in this work, presented for the first time in a seminal article by Caraglio, Baldovin and Stella [4], extends the scope of the Product Space by supplying an explicit representation of its dynamical properties, with the purposes of generating a reliable reproduction of the evolution of the baskets and determining the most relevant factors contributing to their growth. Differently from the approach proposed in [2], the links between the nodes play an active dynamical role and exercise an effective causal influence on the evolution of the whole network. Furthermore, the strength of the links is quantified through the entries of an asymmetric *transfer matrix*, whose structure is directly inspired by Newton's gravitational law (employed in the economic field for the first time by Tinbergen in [5]): specifically, such entries are directly proportional to the time correlation between the historical trading records, which plays the role of an inverse of a physical distance. As will be clarified in the section dedicated to the detailed description of the model, the intrinsically asymmetric nature of the transfer matrix and the embodiment of the empirical correlation allow to effectively solve the inconveniences of the proximity concept.

From a technical standpoint, as already mentioned, the model illustrates the dynamics of the Product Space by means of a system of coupled SDE, each devoted to the evolution of a single trade across the selected period of interest. Theoretical modeling through systems of this kind has been extensively applied in recent years for the description of growth-related issues in several fields, including population and evolutionary dynamics [6], portfolio optimization [7], interface expansion [8] and optimal pinning of vortexes in defected materials [9]. Another notable example in this context, which also directly inspires the approach of [4], is provided by the macro-economic analysis proposed by Bouchaud and Mézard in [10]: in this article, the authors address the problem of the wealth distribution inside a society by means of the formulation of a peculiar system of coupled SDE, each representing a discrete version of the equation that regulates the stochastic diffusion of heat. Along the lines of this last work, each equation of the dynamic model presented in [4] is structured in two distinct sections, respectively devoted to the decoupled component of the evolution of the selected trade

(attributable, for example, to random low-scale events and to the inflationary growth of prices) and to the causal influence exercised on the latter by the remaining part of the network; in particular, the coupling constants which bind the equations consist precisely in the entries of the transfer matrix introduced before. Interestingly, as will be shown in the subsequent chapters, the intertwined structure of the model generates an independent contribution to the total growth of the basket, which can be in principle altered through the tuning of the coupling constants: this feature opens the way to the application of the model for the solution of exploration-exploitation trade-off problems [11–13] related to the commercial activity of the countries, which include the determination of the optimal rate of transfer of resources between different market sectors.

Changing perspective, the main goal of the complexity measures introduced by Teza, Caraglio and Stella in [14] (which represents a conceptual continuation of [4]), extensively analyzed in [15] and applied in this work is the quantitative assessment of the influence exercised by the diversification of the trading baskets on the economic efficiency of a country. The analysis presented in these articles represents the pinnacle of a branch of research started by Hidalgo and Hausmann in [16] and subsequently developed in the works by Pietronero [3, 17–19] and in several others. Despite the dissimilarity between the approaches, an interesting feature shared by all these contributions is the recognition of a mutual dependence between the diversification of the national baskets and the one of the global markets of individual products, which alter each other's quantification in a complex and intertwined manner. In order to address this scenario, with reference to the topological overview presented above, the aforementioned complexity measures supply a static representation of the webs of links connecting individual nodes on each of the two layers to its opposite one, focusing in particular on the distribution of their relative strength and on the interdependence between the two typologies of webs.

As mentioned before, the bare versions of the measures coincide with the celebrated Shannon's entropy function, an extremely flexible object introduced in [20] in the context of communication theory and later applied in adapted forms in many and various scientific fields, including ecology [21], economics [22] and abstract mathematics [23]. Interpreting the relative monetary amounts of the trades related to a given country or to a given product as components of an abstract probability vector, the Shannon's function exhibits a direct proportionality with both their number and their evenness and thus provides a quite natural measure of the diversification of their distribution. In order to properly quantify the mutual dependence between the two kinds of diversification, the bare measures are then subjected to a refinement process through the application of an iterative algorithm, which generates a progressive, effective alteration of the initial distributions of the trades. Due to the smoothness of the Shannon's function, the algorithm presents a fast and safe convergent behavior and provides a solution to some major issues affecting other approaches, like the conceptual ambiguity of the measures presented in the seminal work by Hidalgo and Hausmann [16] and the nonlinear, divergent character of the ones defined by Pietronero in [3].

Despite the generality of the theoretical framework presented above, the econophysics research has concerned, up to the present, only the export activity of the countries. Intuitively, the interest in this field stems from the possibility to employ the composition of the basket exported by a country as an access key to its productive infrastructure [16]. Broadly speaking, the ensemble of capabilities which constitute the latter is quite variegated and comprises both tangible elements (like monetary capital, manpower and raw materials) and intangible ones (like property rights, investment policies and specific labour skills); in particular, the capabilities of the latter kind cannot be effectively spread across the network of countries and are directly responsible for the increasing income disparity typical of today's World [24, 25]. In turn, the richness and the complexity of the productive infrastructure of a country represent determining factors of its global wealth and its economic efficiency, according to the standard economic theory originally formulated by Smith [26] and later developed in several recent works [27, 28].

On the other hand, the import activity presents a much weaker correlation with the productive sector and is related to the wealth of the countries in a conceptually different way. Firstly, considering that the import activity often mainly reflects the actual needs of a society, the analysis of the composition of an import basket provides an estimation of the respective degree of self-sufficiency, a property which is positively correlated with the overall degree of development. In addition, as a consequence of the fierce integration process that has interested the national economies during the last century, both the export and the import activities show relevant repercussions also from the requirements of the global productive system, presenting today an international character: under this circumstance, the quantitative comparison between the distributions of the trading baskets produced by a country provides an assessment of the degree of integration inside such system. In light of these features and of the flexibility of the approach presented in [4] and [14], it is then easy to conclude that the implementation of the methods described above into the analysis of the import activity can potentially supply meaningful information to complement the state of the art of economic complexity, allowing to depict a more complete representation of the nexus between commercial activity and economic growth.

The present work is organized as follows. First of all, Chapter 2 provides an exhaustive representation of the database containing the whole information employed in the analysis. In particular, the description of the bare data organization is followed by a thorough report of the aggregation procedure which subtends the construction of both the export and import bipartite networks; afterwards, a short paragraph depicts the most relevant features pertaining to the hierarchy and the evolution in time of the refined data, which are frequently applied in later sections of the work.

Then, the subsequent couple of chapters is devoted to the technical formulation of, respectively, the dynamic model and the complexity measures presented previously. Regarding Chapter 3, the opening section is completely dedicated to the detailed definition of all the parameters appearing in the equations of the model, describing the role played by each one in the evolution of the trades and the respective macro-

economic interpretation; then, a short section is devoted to the exposition of all the technical and conceptual factors which need to be considered throughout the numerical integration of the model; finally, the calibration procedure of the parameters, which allows to reach an optimal matching degree between the simulations generated by the model and the empirical evolution of the records, is fully reported in the conclusive section of the chapter.

On the other hand, Chapter 4 starts with the formal definition of the bare versions of the complexity measures and continues with the exposition of the conceptual link which relates them with the diversification of the trading baskets; then, the subsequent section presents the iterative algorithm accounting for the refinement of the measures and describes its most relevant mathematical properties.

After the presentation of the models, Chapter 5 is entirely devoted to the presentation of the main results of the analysis of the export bipartite network, mainly following the narrative of the articles by Teza, Caraglio and Stella [4, 14, 15, 29]. The structure of the chapter can be divided, conceptually, into three distinct parts. The first one consists in two sections dedicated, respectively, to the validation of the dynamic model through the reproduction of some key features of the database and to a counterfactual analysis of the growth prospects of a set of relevant countries; then, the second part presents the evaluation of the diversification of the export baskets by means of the refined complexity measures and describes their role as estimators of the economic efficiency of the countries; finally, the last part of the chapter unifies the purposes of the two analytic methods in order to realize a novel dynamical characterization of the countries' network, provided by both the calibrated parameters of the dynamic model and the complexity measures.

Lastly, Chapter 6 supplies an original analysis of the causal relation between the global wealth of the countries and the overall diversification of their commercial activity. After an opening section dedicated to a brief analysis of the import bipartite network, mainly focused on the most relevant analogies and differences which relate it to the export one, the chapter provides an exhaustive representation of the analytic paradigm which allows the evaluation of the dynamical correlation between two quantities, denominated *convergent cross-mapping method* and originally employed in a study on the evolution of complex ecosystems [30]. In particular, the analysis concerns the total income of a country and the diversification of the related export and import baskets, firstly in an ensemble of relevant empirical scenarios and then in a synthetic one generated by means of the dynamic model.

2 Databases and data aggregation

The information concerning the global trade network has been completely extracted from the BACI database, a set of panels compiled annually by the CEPII economic research center [31]. In particular, BACI is the result of a detailed revision [32] of the well-known United Nations COMTRADE database [33], which essentially consists in a series of refining operations (including, for example, the subtraction of transport and insurance costs) having the main purpose of matching the export and the import information. The final database spans a 25-years period between 1995 and 2020 and collects the records of the trades of more than 5000 product categories between more than 200 countries. Each panel of BACI provides a representation of the global market related to a specific year and is organized in the following column structure:

Exporter	**Importer**	**Product Code**	**Value (USD $\cdot$ 10^3)**	**Quantity (t)**

Specifically, the monetary values of the trades, expressed in thousands of US dollars, refer to the specific year of the panel which collects them: in other words, the inflationary growth of prices over time is not intrinsically covered by BACI. Furthermore, the products are organized according to the Harmonized System (HS) [34], a standardized classification method globally implemented in the early 1990s which labels each of them with a 6-digit hierarchical code. As briefly depicted in the following list, moving from the least significant digit towards the most significant one of a given code, larger and more comprehensive categories of products are represented:

- 01----: Live animals;
  - 0101--: Live horses, asses, mules and hinnies;
    - 010121: Purebred breeding horses;
    - ...;
  - 0102--: Live bovine animals;
  - ...;
- ...;
- 49----: Printed books, newspapers and pictures;
  - 4901--: Printed books, including Braille books;
    - 490191: Dictionaries and encyclopedias, and serial instalments thereof;
    - ...;
  - 4902--: Newspapers, journals and periodicals;
  - ....

The peculiar structure of this classification allows to create coarse-grained versions of the database by simply aggregating the product categories which share the first 4 or the first 2 most significant digits. In particular, the analysis presented in this work has been conducted at the 4-digit aggregation level on a total of 1241 product categories, hereafter simply referred to as *products*. Subsequently, further aggregation procedures have been performed in order to collect in a single record either the total export or the total import by a country of a given product, regardless of the external countries standing on the opposite side of the trades. These operations have led to the definition of the trading bipartite networks mentioned in the introduction, which represent the formal topological cornerstones of the whole analysis.

Despite the high precision standard employed in its construction, BACI is currently able to represent only up to the 95% of the whole global market during the period 1995-2020: the lack of some information manifests itself as the occasional absence of some records. These gaps affect both the basic version of the database and (more rarely) the 4-digit coarse-grained one. Since it is unlikely for a country to suddenly stop exporting a whole category of products, the defects at the aggregated level are likely to be attributable to intrinsic issues in the redaction of the original COMTRADE database: confidentiality in the official declarations and the lack of a proper internal organization (in the case of underdeveloped countries) may be proposed as possible causes. Moreover, in some instances the database collects the records from a bunch of different countries in aggregate form up to a certain year, after which the records are presented individually: this is the case of Belgium and Luxembourg (separated from 1999 onwards), Botswana, Namibia, Lesotho, eSwatini and South Africa (from 2000), Serbia and Montenegro (from 2006) and Sudan and South Sudan (from 2012). As a consequence, the time series related to each of these "crippled" countries are all effectively interested by the presence of only gaps before the respective year of separation. In the application of the dynamic model (see Eq. 3.1), treating the gaps as zeros would result in the generation of unmanageable singularities: for this reason, whenever a gap appears, the whole time series which contains it must be discarded. Since this elimination procedure invalidates the study of the countries affected by the presence of a huge number of gaps, the analysis performed through the dynamic model has concerned only the baskets which preserve at least the 80% of their monetary value and at least 20 products after every gap has been ruled out. Of course, this way of proceeding results in the total exclusion of the crippled countries from this branch of the analysis. Overall, the whole elimination procedure has resulted in the exclusion of a total of 111 countries, while 121 is the number of countries considered eligible for the application of the dynamic model; in monetary terms, however, only the 4.3% of the whole global market during the period 1995-2020 has been discarded. Conversely, the analysis of the internal distribution of the baskets through the complexity measures (see Eq. 4.11) is not affected by the occurrence of isolated gaps: thus, the only requirement that needs to be satisfied is the exclusion of the crippled countries whenever the evolution over time of the measures has to be considered in its entirety. Lastly, in some cases the COMTRADE database fails to provide the identity of the exporter or the importer in a given trade. In order to deal with this issue, the redactors of the BACI database have aggregated all these records and have related them to a couple of fictitious countries named "Europe, not else specified" and "Asia, not else specified".

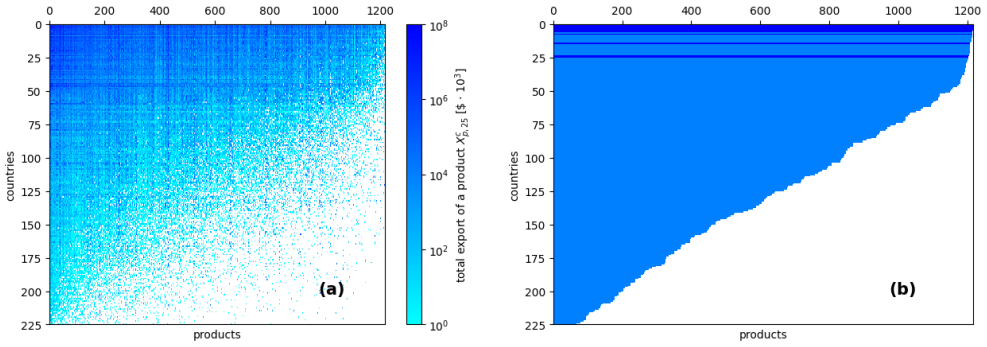

Figure 1: bi-adjacency matrix **(a)** and survival function's diagram **(b)** related to the export bipartite network in year 2020. In panel **(b)**, a set of darker lines highlights the position of the top-10 section of the GDP ranking.

Given the impossibility to trace these records back to the actual countries which have participated in the trades and given their irrelevance inside the European and the Asian markets, both the fictitious countries have been completely ruled out from the analysis.

Since, as already mentioned in the introduction, the first part of the analysis has concerned only the export activity of the countries, the remaining part of this paragraph is devoted to the presentation of some interesting features immediately deducible from the export aggregated database; on the other hand, the analogous features related to its import counterpart will be presented in the final chapter of this work. First of all, an effective way to describe the general structure of the export bipartite network in a given year consists in the construction of the related bi-adjacency matrix, which allows to get an immediate glimpse on the general distribution of the links and to identify the most relevant nodes of both layers. For example, Figure 1a depicts the matrix related to year 2020, putting the nodes in order according to the number of links branching off from them and assigning each of them a color reflecting the monetary value of the respective export. On the other hand, Figure 1b presents a compact form of the same matrix, which plays the role of a survival function's diagram and allows to appreciate the linear growth which characterizes the hierarchy of the links; moreover, a set of darker lines highlights the position of the countries which occupy the top-10 section of the GDP ranking. Following the example of the figure, the export of product p by country c in year n will be hereafter denoted with $X^c_{p,n}$, setting the convention $(n = 0) \leftrightarrow 1995$ and $(n \equiv T = 25) \leftrightarrow 2020$.

Then, Figure 2 shows the evolution of the export basket by United States (USA, panel (a)), China (CHN, panel (b)) and Russia (RUS, panel (c)) at the 2-digit aggregation level, across the period 1995-2020. The products which have been interested by the most relevant export growth (in relative terms) are marked with a thick red line. Otherwise, the wavelength of the color assigned to each product is proportional to the logarithm of the average monetary value of its export during the last five-years period covered by the database. The rainbow effect resulting from this assignment allows to get a glimpse on the degree of stability which has characterized the economy

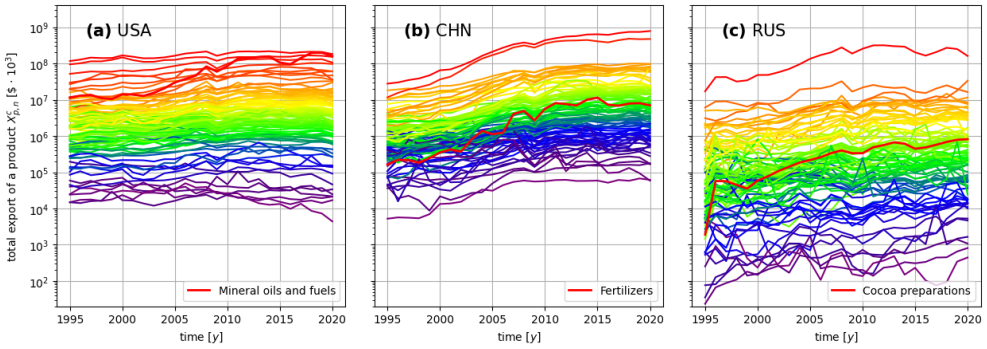

Figure 2: empirical evolution across the period 1995-2020 of the export baskets of United States (USA, **(a)**), China (CHN, **(b)**) and Russia (RUS, **(c)**) at the 2-digit aggregation level. The wavelength of the color assigned to each product reflects its rank (Eq. 2.1) inside the respective basket. For each country, the product interested by the most relevant growth (in relative terms) is marked with a thick red line.

of a country over the selected period: in the example of Figure 2, the economy of the United States appears not to have been affected by any relevant structural change, while China's and Russia's ones have followed a much more turbulent evolution. The hierarchy of the products in the baskets can be described through the definition of proper rankings, which slowly evolve in time: denoting with N_n^c the number of products contained in the basket of country c in year n, the rank of product p is defined as

$$x_p^c \equiv \frac{1}{5} \sum_{n=0}^{T} \frac{X_{p,n}^c}{\Omega_n^c} \quad , \tag{2.1}$$

with

$$\Omega_n^{ex,c} \equiv \sum_{p=0}^{N_n^c} X_{p,n}^c \quad .$$

As can be easily imagined, the reproduction of the final rankings of the products, starting from the real initial conditions of year 1995, will represent an important test of the validity of the dynamic model. Another property which can be easily noted from Figure 2 is the exponential growth which generally interests the baskets in their totality (the China's one in somewhat relevant way). In order to properly compare the evolution of the baskets from different countries, it is convenient to characterize their growth from an intensive perspective by considering the average logarithmic returns over the full time range $[0, T]$:

$$\lambda_T^c \equiv \frac{1}{N_n^c \cdot T} \sum_{p=0}^{N_n^c} \log \left(\frac{X_{p,T}^c}{X_{p,0}^c} \right) \quad ; \tag{2.2}$$

this quantity will be hereafter referred to as the *empirical growth* of a country and will play an important role in the calibration of the model (see Section 3.3).

Lastly, the online archive of the World Bank Group [35] has supplied all the information concerning both the inflation of the prices and the gross domestic product (GDP) at the country level. In particular, the annual inflation percentage rate has been estimated through the annual GDP deflator related to the United States (the monetary value of the trades in BACI being always expressed in US dollars): since it covers the whole range of goods and services produced in the economy and not only the limited consumers' commodities basket, the GDP deflator is commonly deemed as a more comprehensive inflation measure than the frequently employed consumer price index (CPI). In the case of inflation, the records of year n will be indicated with I_n.

3 Dynamic model of the trades

This chapter is entirely devoted to a thorough description of the dynamic model by Caraglio, Baldovin and Stella, presented in the seminal work [4] with the purpose of providing a reliable representation of the dynamical properties of the Product Space. More precisely, the original aim of the model was the description of the evolution over time of the global market from the export perspective, aggregating the trades between the countries at the global level. Afterwards, the scope of the model was extended in the works by Teza, Caraglio and Stella [14,29], where the effectiveness of its application to the export activity at the country level was proven. In the spirit of all these works, the discussion presented in this chapter focuses on the contextualization of the model in the export scenario, describing the conceptual link between the parameters appearing in the equations and the productive infrastructure of the countries; on the other hand, the application of the model to the import activity will be presented in Chapter 6.

The upcoming section presents the general structure of the equations of the model and the basic principles which subtend the formulation of all their components; subsequently, a short section is devoted to the exposition of the scheme adopted for the numerical integration of the model, focusing on the technical and conceptual factors which justify its selection; finally, the procedure adopted for the calibration of the parameters is reported in details in the last section of the chapter.

3.1 Structure of the model

From the technical point of view, the model consists in a system of coupled stochastic differential equations (SDE). Since the yearly variations of the records are the average result of fluctuations which take place during much shorter periods, the evolution over time of the export of product p by country c is described by the continuous function $X_p^c(t)$, with time expressed in years. In this perspective, the quantities $X_{p,n}^c$, introduced in Chapter 2, define the historical discrete representation in 26 points of the evolution; in order to reproduce it, the initial condition of the system must be set to $X_p^c(0) = X_{p,0}^c$, $\forall p$ (the starting point $t = 0$ coinciding with the end of year 1995). Each equation of the system, related to a particular product exported by country c, takes the following form:

$$\frac{\partial X_p^c(t)}{\partial t} = [\mu^c(t) + \eta_p^c(t)]X_p^c(t) + \sum_{p' \neq p} [J_{pp'}^c X_{p'}^c(t) - J_{p'p}^c X_p^c(t)] \quad . \tag{3.1}$$

As can be immediately noticed, two different parts can be distinguished in the right hand side of Eq. 3.1. The term proportional to $X_p^c(t)$ describes, in terms of a generalized Brownian motion, the decoupled component of the evolution of the export; on the other hand, the second part takes the form of a master equation and defines a mechanism of transfer of resources accounting for the collective influence exercised by the whole basket. As will be demonstrated in Section 3.2, Eq. 3.1 can be easily recast into the form of a general Itō equation. The next paragraphs are devoted to the detailed description of the role played by the parameters of the model and of the technical derivation of their properties.

3.1.1 Correlation matrix

As will be explained later in this section, both the coupled and the uncoupled component of the evolution of the exports by country c are heavily influenced by the presence of a correlation matrix, denoted with $\mathbf{C^c}$, which quantifies the strength of the connection between different products. In particular, each entry of this matrix is calculated as a formal time correlation between the historical variations of a pair of exports. The first step of the construction of $\mathbf{C^c}$ consists in the precise definition of such variations. Considering the exponential growth which globally interests every basket, it is convenient to define them in terms of yearly logarithmic returns: in particular, the return at time n related to product p is calculated as

$$R^c_{p,n} \equiv \log\left(\frac{X^c_{p,n}}{X^c_{p,n-1}}\right) \quad . \tag{3.2}$$

Since the interest lies in the deviation of such returns from the respective means, it is useful to immediately implement the corresponding normalized quantities:

$$r^c_{p,n} \equiv \frac{R^c_{p,n} - \langle R^c_{p,n}\rangle}{\sqrt{\mathrm{Var}(R^c_{p,n})}} = \frac{R^c_{p,n} - \langle R^c_{p,n}\rangle}{\sqrt{\langle (R^c_{p,n})^2\rangle - (\langle R^c_{p,n}\rangle)^2}} \quad , \tag{3.3}$$

where each mean is calculated over the whole period as

$$\langle R^c_{p,n}\rangle = \frac{1}{T}\sum_{n=1}^{T} R^c_{p,n} \quad . \tag{3.4}$$

Finally, the entries of the matrix $\mathbf{C^c}$ are obtained as time correlations between the normalized returns:

$$C^c_{p,p'} = \frac{1}{T}\sum_{n=1}^{T} r^c_{p,n} r^c_{p',n} \quad . \tag{3.5}$$

The structure of the correlation matrix allows to properly simulate the intertwined nature of the export baskets. Specifically, given its empirical character, the matrix manages to keep track of the connection between exports which are evidently distant from the productive point of view. As an example, it is plausible to assume that an increase in the export of building bricks, presumably caused by an increase of the internal demand for new buildings and by a consequent increase of their production, is positively correlated to the production and export of refrigerators and freezers (usually bought only once when moving to a new home). In agreement with this situation, the related entry of the global correlation matrix has an above-average value equal to 0.77 ($\langle C^{WRD}_{pp'}\rangle = 0.44$). On the other hand, the proximity measures which incorporate only infrastructural and productive similarities between the exports (like the one employed in [16]) usually fail to catch this and other similar connections.

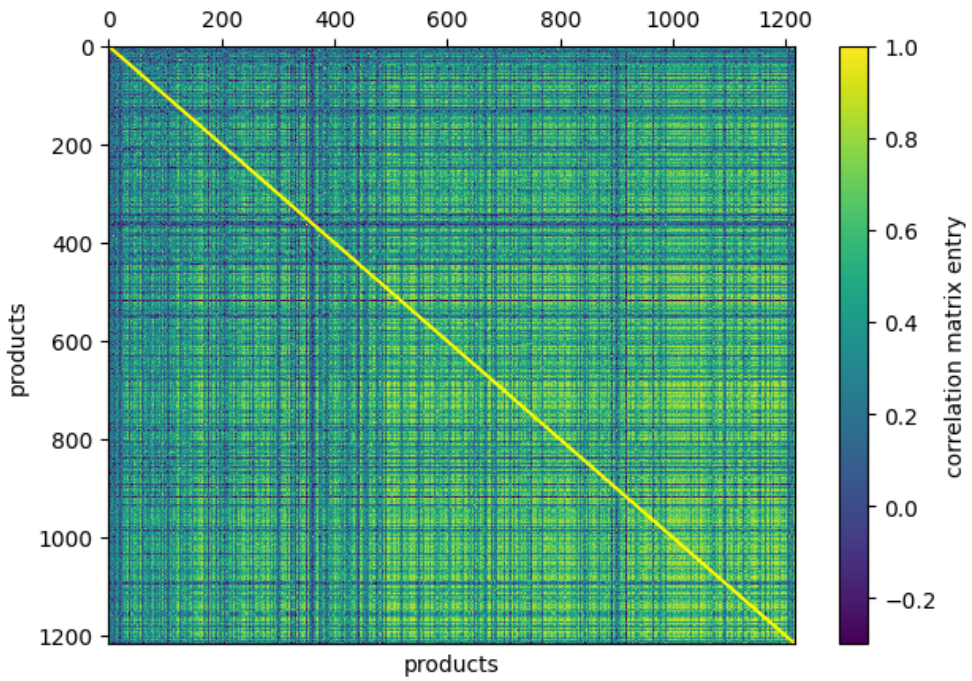

Figure 3: empirical correlation matrix related to the exports of the countries aggregated at the global level. The products are distributed according to their global ranks.

Figure 3 depicts the structure of the correlation matrix related to the global aggregation of all the exports. On both axes, the products are distributed according to their global ranks. This feature allows to observe the correlation structures which characterize in a widespread manner the global database and which manifest themselves as clear thick bands; moreover, the dominant presence of positive bands in the high-rank regions allows to perceive the strong influence typically exercised by the most exported products on vast portions of the global market. As a last remark, it should be noted that, given its symmetric structure, the correlation matrix does not carry any causal information concerning the coupled evolution of the pairs of exports.

3.1.2 Drift term

The first coefficient in the right hand side of Eq. 3.1, $\mu^c(t)$, is a deterministic drift, accounting for the average, independent exponential growth rate of the exports of country c. In principle, a more sophisticated version of the model could involve the assignment of a different drift coefficient to each single export; however, such a modification, despite being attractive from the theoretical point of view, would implicate a dramatic increase of the number of parameters defining the model, leading to a drastic complication of the calibration procedure and compromising the minimalist

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003
$\boldsymbol{I_n}$	2.10	1.83	1.79	1.13	1.44	2.24	2.19	1.58	1.86
Year	2004	2005	2006	2007	2008	2009	2010	2011	2012
$\boldsymbol{I_n}$	2.69	3.11	3.03	2.69	1.95	0.76	1.17	2.09	1.92
Year	2013	2014	2015	2016	2017	2018	2019	2020	
$\boldsymbol{I_n}$	1.75	1.85	0.95	1.05	1.88	2.40	1.79	1.17	

Table 1: annual percentage inflation rate of the US dollar across the period 1995-2020.

spirit which inspired from the beginning the whole analysis. Moreover, a quick look at the plots shown in Figure 2 reveals that, ignoring noise effects, the logarithms of the exports grow at a pretty similar rate, justifying the implementation of a common average drift coefficient. The inclusion of a time dependence into its definition satisfies the necessity to describe the inflationary growth of prices over time, which is not covered by BACI. For this reason, the coefficient is subdivided into a constant parameter, accounting for the financial investments in the productive sector, and a time dependent term, describing the inflation of the US dollar: $\mu^c(t) = \overline{\mu}^c + I(t)$. The function $I(t)$ takes a simple annual step-wise form (depicted in Figure 4) and satisfies the condition $I(t) \equiv I_n$ for $(n-1) \leq t < n$ $(n = 1, ..., T)$, where the empirical values I_n are directly extracted from the World Bank archive and are shown in Table 1.

3.1.3 Noise term

The second coefficient appearing in the model, $\eta_p^c(t)$, is a colored Gaussian noise term with zero average and unit variance, accounting for random variations of the exports due to low-scale events in the productive sector. Figure 2 shows that the volatility of the evolution, in contrast to the exponential growth, can change significantly inside a given basket and seems to be less pronounced in the most exported sections. On one hand, this property justifies the inclusion of a different noise coefficient for each export; on the other hand, it does not necessarily imply that the stochastic component of the evolution of a single export should be completely independent from all the others. Indeed, it is natural to suppose the intertwined nature of the baskets, described by means of the correlation matrix, to exercise a considerable influence on the evolution of the exports also at the shortest time scales. In addition, the occurrence of a persistent low-scale event can in principle stimulate or dampen the evolution of a single export (at least for a limited amount of time), independently of the behavior of the rest of the basket. These considerations justify the colored nature of the noise term, which can be realized through the implementation of both a time correlation and a correlation between different exports. In particular, the noise term is required to satisfy the following properties:

$$\left\langle \eta_p^c(t) \right\rangle = 0 \qquad \forall p, t \tag{3.6}$$

$$\left\langle \eta_p^c(t) \eta_{p'}^c(t') \right\rangle = C_{pp'}^c \frac{(\sigma^c)^2}{\tau^c} \mathrm{e}^{-|t-t'|/\tau^c} \qquad \forall p, p', t, t' \quad , \tag{3.7}$$

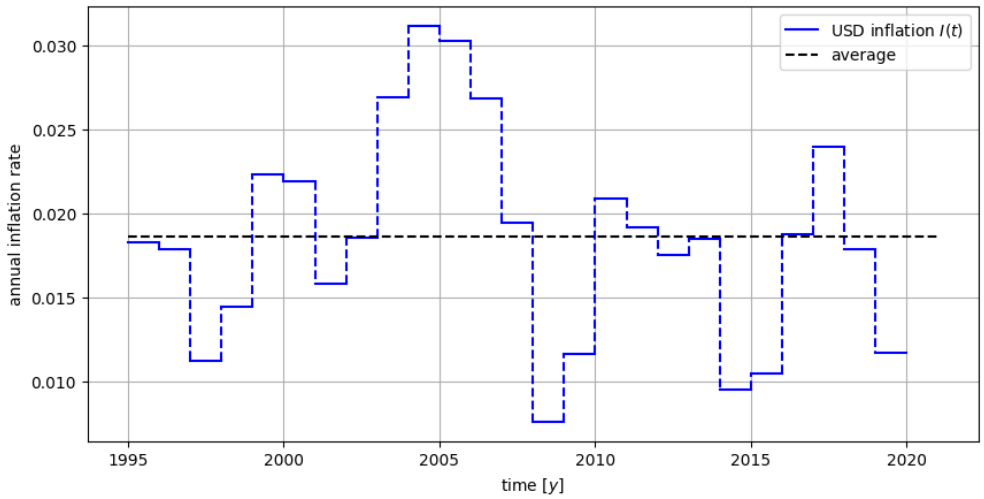

Figure 4: evolution of the US dollar inflation function $I(t)$ across the period 1995-2020.

where the means are extracted from different realizations. These equations enrich the dynamic model with two further parameters, related to the time correlation of the noise term: in particular, σ^c describes the overall extent of the random fluctuations in the evolution of the basket, while τ^c quantifies the duration of possible periods characterized by persistent growth or recession; on the other hand, the correlation between different exports is described once again by the correlation matrix. The validity of these choices will be ultimately tested by checking the ability of the model to reproduce some important features of the export database.

In order to implement the properties shown in Eq. 3.7, it is necessary to generate, first of all, a completely uncorrelated Gaussian noise term $\vec{\xi^c}(t)$ with zero average and unit variance:

$$\left\langle \xi_p^c(t) \right\rangle = 0 \qquad \forall p, t \tag{3.8}$$

$$\left\langle \xi_p^c(t) \xi_{p'}^c(t') \right\rangle = \delta_{pp'} \delta(t - t') \qquad \forall p, p', t, t' \quad . \tag{3.9}$$

At this point, the LDL^T version of the Cholesky decomposition has to be performed on the correlation matrix. This particular operation allows to decompose a generic symmetric square matrix (not necessarily positive definite) into the product of a lower triangular matrix ($\mathbf{L^c}$), a diagonal one with non-negative entries ($\mathbf{D^c}$) and the transpose of the former. The main advantage of this version of the decomposition, with respect to the classical LL^T one, consists in the possibility to avoid the calculation of square roots, a step that often leads to severe issues when large matrices with small entries are involved. The entries of the matrices involved are obtained through the

following recursive algorithm (for the details concerning its construction, see, for instance, [36]):

$$D^c_{ii} = C^c_{ii} - \sum_{k=0}^{i-1} D^c_{kk} L^c_{ik} \qquad \forall i = 0, ..., \dim\mathbf{C^c} \tag{3.10}$$

$$L^c_{ji} = \frac{C^c_{ij} - \sum_{k=0}^{i-1} D^c_{kk} L^c_{ik} L^c_{jk}}{D^c_{ii}} \qquad \forall i = 0, ..., \dim\mathbf{C^c}, \quad j = i+1, ..., \dim\mathbf{C^c} \quad ; \tag{3.11}$$

as shown, the calculation of a given entry of $\mathbf{D^c}$ is followed by the calculation of all the entries of $\mathbf{L^c}$ on the same column. The practical implementation of this algorithm through a code usually leads to inconveniences due to approximation errors. Eq. 3.10 clearly shows that the sequence $\{D^c_{ii}\}_{i=0,...,\dim\mathbf{C^c}}$ rapidly decreases as the order of the rows increases (although not necessarily in a monotonic way), eventually making the entries cross the long double precision limit; this circumstance, as Eq. 3.11 shows, causes in turn the divergence of the entries of $\mathbf{L^c}$. The crossing of the limit always takes place after just a few tens of iterations, while $\dim\mathbf{C^c}$ is often of the order of a thousand. Since the values of the subsequent entries are in any case negligible, it is possible to achieve the decomposition at an excellent approximation level by setting them to 0: after a validity check, the mean difference between the entries of $\mathbf{C^c}$ obtained in this way and the ones of the product $\mathbf{L^c D^c} (\mathbf{L^c})^T$ is always found to be of the order of 10^{-12} or lower. After performing the decomposition, it is easy to verify that the noise term $\vec{\zeta}^c(t) = \mathbf{L^c} (\mathbf{D^c})^{1/2} \vec{\xi}^c(t)$ satisfies the desired correlation property between different exports:

$$\left\langle \zeta^c_p(t) \zeta^c_{p'}(t') \right\rangle = C^c_{pp'} \delta(t - t') \qquad \forall p, p', t, t' \qquad . \tag{3.12}$$

On the other hand, the required time correlation property can be obtained, at the infinitesimal level, through a proper definition of the evolution of the colored noise term: in particular, the choice fell on the expression

$$\eta^c_p(t) = \rho^c \eta^c_p(t - \mathrm{d}t) + \frac{\sigma^c}{\sqrt{\tau^c}} \sqrt{1 - (\rho^c)^2} \zeta^c_p(t) \qquad \text{with} \quad \rho^c = \mathrm{e}^{-\mathrm{d}t/\tau^c} \qquad , \tag{3.13}$$

with the initial condition

$$\eta^c_p(t_0) = \frac{\sigma^c}{\sqrt{\tau^c}} \zeta^c_p(t_0) \qquad . \tag{3.14}$$

Selecting the particular case $t = t_0 + \mathrm{d}t$ and $t' = t_0$, a straightforward calculation shows that Eq. 3.7 is finally recovered:

$$\left\langle \eta^c_p(t_0 + \mathrm{d}t) \eta^c_{p'}(t_0) \right\rangle =$$

$$= \frac{\sigma^c}{\sqrt{\tau^c}} \rho^c \left\langle \eta^c_p(t_0 + \mathrm{d}t - \mathrm{d}t) \zeta^c_{p'}(t_0) \right\rangle + \frac{\sigma^c}{\sqrt{\tau^c}} \sqrt{1 - (\rho^c)^2} \left\langle \zeta^c_p(t_0 + \mathrm{d}t) \zeta^c_{p'}(t_0) \right\rangle =$$

$$= \frac{(\sigma^c)^2}{\tau^c}\rho^c \left\langle \zeta_p^c(t_0)\zeta_{p'}^c(t_0)\right\rangle + \frac{\sigma^c}{\sqrt{\tau^c}}\sqrt{1-(\rho^c)^2}\left\langle \zeta_p^c(t_0+\mathrm{d}t)\zeta_{p'}^c(t_0)\right\rangle =$$

$$= C_{pp'}^c \frac{(\sigma^c)^2}{\tau^c}\rho^c + C_{pp'}^c \frac{\sigma^c}{\sqrt{\tau^c}}\sqrt{1-(\rho^c)^2}\underbrace{\delta(\mathrm{d}t)}_{=0} = C_{pp'}^c \frac{(\sigma^c)^2}{\tau^c}\rho^c \qquad ;$$

exploiting the correlation's invariance under time translations, the validity of this result can be immediately extended to any couple of infinitesimally close instants. The exact fulfillment, through the choice of Eqs. 3.13 and 3.14, of Eq. 3.7 only at the infinitesimal level is still satisfactory. Indeed, in order to perform the integration of the model (Section 3.2), the period of interest must be subdivided into a huge number of steps $\{t_n\}_{n=0,\dots,N}$, with $\Delta t \equiv t_n - t_{n-1} \ll 1$: as a matter of fact, such a property can be practically imposed only between consecutive steps and is then only approximately respected when longer periods are considered.

3.1.4 Transfer matrix

The coupled component of the evolution of the exports is determined by the properties of the matrix $\mathbf{J^c}$, whose task is to properly quantify the rate of transfer of resources between different exports. In light of the complexity of the productive sector of a country, a phenomenon of this kind may stem from a variety of different factors, including the similarity in the organization of different productive chains or the implementation of particular investment policies and regulations.

The first condition that should be satisfied by $\mathbf{J^c}$ is the absence of a reflection symmetry with respect to its diagonal, in order to properly describe the asymmetric nature of the causal relationship between any pair of exports: $J_{pp'}^c \neq J_{p'p}^c$. For example, this feature is immediately noticeable in the case of the pair of products considered previously: while an increase of the export of refrigerators could stem from an increase of the export of bricks, the opposite circumstance does not occur.

Furthermore, the transfer of resources should be set in order to reproduce, as a result of a numerical simulation starting from the real initial distribution of the exports, the empirical ranking defined in Eq. 2.1, denoted with $\vec{x}^c$. The implications of this task on the structure of $\mathbf{J^c}$ can be addressed, as a first step, by recasting the coupled part of Eq. 3.1 into a different form (obviously, the decoupled part does not play any role in the reproduction of $\vec{x}^c$):

$$\left[\frac{\partial X_p^c(t)}{\partial t}\right]^{cp} = \sum_{p'\neq p}\left[J_{pp'}^c X_{p'}^c(t) - J_{p'p}^c X_p^c(t)\right] = \sum_{p'} A_{pp'}^c X_{p'}^c \qquad , \tag{3.15}$$

where $A_{pp'}^c \equiv J_{pp'}^c$ for $p \neq p'$ and $A_{pp}^c \equiv -\sum_{p'\neq p} J_{p'p}^c$. Eq. 3.15 is a system of linear differential equation with constant coefficients, whose forward flux can be generated, for example, by the time dependent exponential matrix $\mathrm{e}^{t\mathbf{A^c}}$. As proven in [37], if $\mathbf{A^c}$ has non negative off-diagonal entries (a property shared with $\mathbf{J^c}$) and if it is only composed by columns with zero sum (true by construction), then the matrix $\mathrm{e}^{t_n\mathbf{A^c}}$ (for any $t_n > 0$) describes the evolution of an ergodic Markov process. As a consequence, the repeated application of $\mathrm{e}^{t_n\mathbf{A^c}}$ to a probability vector $\vec{p}$ (such that $\sum_i p_i = 1$) converges to the unique (non degenerate) stationary state of the process. This state,

being a generator of the unitary eigenspace of $e^{t\mathbf{A^c}}$, also generates the kernel of $\mathbf{A^c}$. In order to achieve the task expressed above, the role of stationary state should be played by the empirical ranking $\vec{x}^c$ and the simplest way to realize the convergence of the (normalized) exports consists in imposing the conditions

$$J^c_{pp'} \geq 0 \quad \forall p, p' \mid p \neq p' \qquad \text{and} \qquad J^c_{pp'} \propto x^c_p \quad \forall p \qquad ,$$

where the latter makes $\vec{x}^c$ generate the kernel of $\mathbf{A^c}$ and also fulfills the asymmetry of $\mathbf{J^c}$.

As a final condition, the transfer of resources should obviously reflect the strength of the connection which links the products involved. Similarly to the case of the noise term, this task is accomplished through the implementation of the correlation matrix into the definition of the transfer one, whose entries take the final form

$$J^c_{pp'} = G^c x^c_p |C^c_{pp'}| \qquad ; \tag{3.16}$$

the introduction of the new parameter G^c allows to tune the extent of the transfer mechanism inside the respective basket. The appeal of 3.16 lies in its affinity with the well-known Newton's gravitational law. Interpreting the entry $J^c_{pp'}$ from the physical perspective, the normalized "charge" x^c_p generates, in an average sense, a counterpart of the gravitational field: then, through the coupling with the external charges represented by the exports, the entry produces a "force" whose magnitude is directly proportional to the constant G^c, the charges themselves and the squared inverse "distance" $|C^c_{pp'}|$ between the products.

3.2 Integration of the model

Given their complexity, an exact analytical integration of the equations of the model is clearly out of range: this circumstance implies the necessity of the adoption of a numerical integration scheme for the generation of faithful approximate solutions. During the last decades, different integration schemes for systems of SDE have been proposed and exhaustively analyzed. In the selection of the scheme, the most relevant criterion of choice concerns the convergence of the numerical solutions, depending on the context of analysis. Since a considerable part of the dynamical analysis presented in the subsequent chapters involves the execution of huge sets of simulations, the quality of the reproduction of the average behavior of the exports generated by Eq. 3.1 should be optimized, while the accuracy of the reproduction of single particular realizations can be overlooked. Based on these considerations, the choice fell on a second order Runge-Kutta scheme adapted for systems of SDE, introduced in [38] and depicted in Eq. 3.18. The numerical solutions of this scheme present both a strong convergence and a weak convergence of order 1 and guarantee an optimal trade-off between quality of reproduction and time required for the simulations.

In order to implement the scheme, it is necessary to recast the equations of the model into a more suitable form. First of all, by defining a Brownian motion from each component of the noise term $\vec{\zeta}^c(t)$,

$$\mathrm{d}W_p^c \equiv \zeta_p^c(t)\sqrt{\mathrm{d}t} \qquad ,$$

the evolution of the colored noise term $\vec{\eta}^c(t)$ can be reformulated in the following way:

$$\eta_p^c = \rho^c \eta_p^c(t-\mathrm{d}t) + \frac{\sigma^c}{\sqrt{\tau^c}}\sqrt{1-(\rho^c)^2}\zeta_p^c(t)\frac{\mathrm{d}W_p^c}{\sqrt{\mathrm{d}t}} \qquad .$$

Then, by implementing the above expression and separating the deterministic component from the stochastic one, it is possible to recast the differential version of Eq. 3.1 into the form of a general Itō equation:

$$\begin{aligned}\mathrm{d}X_p^c = \left\{ \overline{\mu}^c + I(t) + \rho^c \eta_p^c(t-\mathrm{d}t) + \sum_{p'\neq p}\left[J_{pp'}^c \frac{X_{p'}^c(t)}{X_p^c(t)} - J_{p'p}^c \right] \right\} X_p^c(t)\mathrm{d}t + \\ + \frac{\sigma^c}{\sqrt{\tau^c}}\sqrt{1-(\rho^c)^2}\sqrt{\mathrm{d}t}X_p^c(t)\mathrm{d}W_p^c \equiv \\ \equiv a_p^c[t,\vec{X}^c(t)]\mathrm{d}t + b_p^c[t,\vec{X}^c(t)]\mathrm{d}W_p^c \qquad . \end{aligned} \tag{3.17}$$

Finally, denoting with $\{\vec{w}_n^c\}_{n=0,\dots,N}$ the discrete representations of the functions $\vec{X}^c(t)$ in the set of N instants $\{t_n\}_{n=0,\dots,N}$, the Runge-Kutta scheme

$$\begin{cases} \vec{w}_0^c = \vec{X}^c(0) \\ \vec{K}_{1,n}^c = \vec{a}^c(t_n,\vec{w}_n^c)\Delta t + \mathbf{B_-^c}\,\vec{b}(t_n,\vec{w}_n^c) \\ \vec{K}_{2,n}^c = \vec{a}^c(t_{n+1},\vec{w}_n^c + \vec{K}_{1,n}^c)\Delta t + \mathbf{B_+^c}\,\vec{b}^c(t_{n+1},\vec{w}_n^c + \vec{K}_{1,n}^c) \\ \vec{w}_{n+1}^c = \vec{w}_n^c + \frac{\vec{K}_{1,n}^c + \vec{K}_{2,n}^c}{2} \end{cases} , \tag{3.18}$$

where $\Delta t \equiv t_n - t_{n-1}$, can be applied to the model. In the above equations, the diagonal matrices $\mathbf{B_-^c}$ and $\mathbf{B_+^c}$ contain the development of the Brownian motion associated with the process:

$$\begin{aligned}(B_-^c)_{pp'} &= \delta_{pp'}(\Delta W_{p,n}^c - S_{p,n}^c\sqrt{\Delta t}) \\ (B_+^c)_{pp'} &= \delta_{pp'}(\Delta W_{p,n}^c + S_{p,n}^c\sqrt{\Delta t}) \qquad ,\end{aligned}$$

where $\Delta W_{p,n}^c \equiv \zeta_{p,n}^c\sqrt{\Delta t}$ and $\{S_{p,n}^c\}$ is a set of numbers assuming the values 1 or -1, both chosen with a 50% probability. As proved in [38], the scheme in Eq. 3.18 can be implemented to reproduce both the Itō's and the Stratonovich's interpretations of a system of SDE (in the latter case, it is sufficient to set $S_{p,n}^c = 0$, $\forall p,n$),

providing solutions with identical convergence properties. However, if the condition $\Delta t \ll \tau^c$ is fulfilled, the outcome of each simulation is independent of the choice of the interpretation.

3.3 Calibration of the model

The calibration procedure described in this section builds a connection between the dynamic model and the export database and aims to determine the values of the parameters that optimally reproduce the empirical evolution of the exports.

As a first step, it is useful to reformulate the model in a different way. Dividing Eq. 3.1 by $X_p^c(t)$ and integrating in the generic interval $[n_1, n_2]$, the following relation is obtained:

$$\log \frac{X_p^c(n_2)}{X_p^c(n_1)} = \int_{n_1}^{n_2} \left(\overline{\mu}^c + I(t) + \eta_p^c(t)\right) \, \mathrm{d}t + \int_{n_1}^{n_2} \sum_{p' \neq p} \left(G^c x_p^c |C_{pp'}^c| \frac{X_{p'}^c(t)}{X_p^c(t)} - G^c x_{p'}^c |C_{p'p}^c| \right) \, \mathrm{d}t \, . \tag{3.19}$$

Then, by defining the functions

$$f_p^c(n_1, n_2) \equiv \frac{1}{n_2 - n_1} \left(\log \frac{X_p^c(n_2)}{X_p^c(n_1)} - \int_{n_1}^{n_2} I(t) \, \mathrm{d}t \right) \tag{3.20}$$

and

$$g_p^c(n_1, n_2) \equiv \frac{1}{n_2 - n_1} \int_{n_1}^{n_2} \sum_{p' \neq p} \left(x_p^c |C_{pp'}^c| \frac{X_p^c(t)}{X_{p'}^c(t)} - x_{p'}^c |C_{p'p}^c| \right) \mathrm{d}t \quad , \tag{3.21}$$

Eq. 3.19 can be rewritten in a more compact form:

$$f_p^c(n_1, n_2) = G^c g_p^c(n_1, n_2) + \overline{\mu}^c + \frac{1}{n_2 - n_1} \int_{n_1}^{n_2} \eta_p^c(t) \, \mathrm{d}t \quad . \tag{3.22}$$

The subsequent paragraphs will show how the evaluation of all the values $f_p^c(n_1, n_2)$ and $g_p^c(n_1, n_2)$ $(\forall p, n_1, n_2 \mid 0 \leq n_1 \leq (T-1),\ 1 \leq n_2 \leq T)$, performed through the implementation of the empirical records, leads to the estimation of the whole ensemble of parameters. To this purpose, given the discrete nature of the database, the integral of the export functions in Eq. 3.21 has to be approximated with a discrete summation:

$$\int_{n_1}^{n_2} \frac{X_p^c(t)}{X_{p'}^c(t)} \, \mathrm{d}t \approx \sum_{n=n_1}^{n_2-1} \frac{1}{2} \left(\frac{X_{p,n}^c}{X_{p',n}^c} + \frac{X_{p,n+1}^c}{X_{p',n+1}^c} \right)$$

3.3.1 Calibration of the transfer parameter

Noting that the transfer parameter G^c plays the role of the slope in Eq. 3.22, its estimation can be in principle obtained through a linear regression of all the pairs of values $\{(f, g)\}$. However, as shown in Figure 5 in the case of United States and Zambia, a significant amount of these pairs always strongly deviates from linearity, hindering

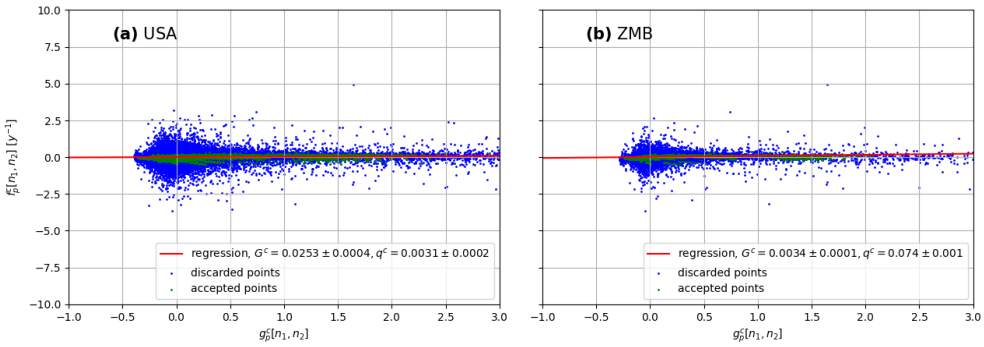

Figure 5: calibration of G^c in the case of United States (USA, **(a)**) and Zambia (ZMB, **(b)**). The pairs of values {f,g} (Eqs. 3.20 and 3.21) subjected to the linear regressions are marked with green points.

the calibration procedure (for several countries, an unacceptable negative value of G^c is obtained). From the perspective of the model, any deviation from a perfect linear dependence with slope G^c and intercept $\overline{\mu}^c$ has to be attributed to the influence of the noise term. Assuming that the time correlation of the latter is not too strong, the model also predicts the tendency by the pairs of points to approach linearity as longer time intervals $[n_1, n_2]$ are considered. For this reason, the calibration procedure has considered in every case only the set of pairs $\{(f, g)\}$ related to a time interval longer than 15 years: this value has been chosen in order to both limit the influence of the noise term and, at the same time, guarantee the presence of a sufficiently high number of points. Figure 5 depicts the distinction between accepted and discarded pairs $\{(f, g)\}$ and allows to appreciate the ability of the model to identify, regardless of the degree of development of the country considered, a subset of points in better agreement with the linear behavior. The value of the intercept resulting from the regression provides a first, raw estimation of $\overline{\mu}^c$; a more reliable procedure of calibration of the latter, which employs the whole information provided by the database, will be described in the final paragraph of this section.

3.3.2 Calibration of the noise parameters

The estimation of the noise parameters σ^c and τ^c relies on a reformulation of the relation which binds the pairs $\{(f, g)\}$. Evaluating Eq. 3.22 in the interval $[0, n]$ and considering the variance with respect to the products of both terms, the following relation is found:

$$\mathrm{Var}\left[\int_0^n \eta_p^c(t)\ \mathrm{d}t\right] = n^2 \mathrm{Var}\left[f_p^c(0,n) - G^c g_p^c(0,n)\right] \quad . \tag{3.23}$$

The next steps aim to express the right hand side of this equation, which can be calculated from the empirical records, as a function of σ^c and τ^c. First of all, recalling a common statistical relation, the left hand side can be expressed in terms of proper

expected values:

$$\mathrm{Var}\left[\int_0^n \eta_p^c(t)\ \mathrm{d}t\right] = \mathbb{E}\left[\left(\int_0^n \eta_p^c(t)\ \mathrm{d}t\right)^2\right] - \left(\mathbb{E}\left[\int_0^n \eta_p^c(t)\ \mathrm{d}t\right]\right)^2 \approx$$
$$\approx \mathbb{E}\left[\left(\int_0^n \eta_p^c(t)\ \mathrm{d}t\right)^2\right] \quad . \tag{3.24}$$

The approximation consisting in neglecting the squared average integral is not obvious and requires some explanation. As a matter of fact, in the single realization context considered in Eq. 3.23, such a quantity cannot be set exactly equal to 0, because the cross-correlation property of the noise term (Eq. 3.7) prevents the derivation, from Eq. 3.6, of the key fixed-time property $\mathbb{E}\left[\eta_p^c(t)\right] \equiv 0$, which should be valid at least in the ideal limit of infinite products (on the other hand, due to the interchangeability of the extractions of the means, the same property is certainly deducible from Eq. 3.6 in an average sense: $\langle \mathbb{E}\left[\eta_p^c(t)\right]\rangle \equiv 0$). However, by assuming that the correlation of the noise term with respect to the products is not too strong, it is still possible to employ the approximation $\mathbb{E}\left[\eta_p^c(t)\right] \approx 0$, which renders the quantity $\mathbb{E}\left[\int_0^n \eta_p^c(t)\ \mathrm{d}t\right]$ negligible: as a consequence, the square of the latter turns out to be completely irrelevant in the computation of the right hand side of Eq. 3.24 and can therefore be discarded.

Developing, on the other hand, the expected value of the squared integral:

$$\mathbb{E}\left[\left(\int_0^n \eta_p^c(t)\ \mathrm{d}t\right)^2\right] = \int_0^n \int_0^n \mathbb{E}\left[\eta_p^c(t)\eta_p^c(t')\right]\ \mathrm{d}t\ \mathrm{d}t' \approx \frac{(\sigma^c)^2}{\tau^c}\int_0^n \int_0^n \mathrm{e}^{-\frac{|t-t'|}{\tau^c}}\ \mathrm{d}t\ \mathrm{d}t' \quad , \tag{3.25}$$

where the last step is induced, on average, by Eq. 3.7. It should be noted that this approximation is much less relevant than the one adopted in Eq. 3.24: indeed, due to the absence of any 4-fields composed correlation ($\langle \eta_p^c(t)\eta_{p'}^c(t')\eta_{p''}^c(t'')\eta_{p'''}^c(t''')\rangle = 0$), in this case it is possible to state the stochastic convergence (in the ideal limit of infinite products) of the quantity $\mathbb{E}\left[\eta_p^c(t)\eta_p^c(t')\right]$ towards the right hand side of Eq. 3.7 (with $C_{pp}^c = 1$). Therefore, implementing the change of variables $t'' \equiv t - t'$:

$$\mathbb{E}\left[\left(\int_0^n \eta_p^c(t)\ \mathrm{d}t\right)^2\right] \approx \frac{(\sigma^c)^2}{\tau^c}\int_0^n \mathrm{d}t' \int_{-t'}^{n-t'} \mathrm{e}^{-\frac{|t''|}{\tau^c}}\ \mathrm{d}t'' =$$
$$= \frac{(\sigma^c)^2}{\tau^c}\int_0^n \mathrm{d}t' \left(\int_{-t'}^0 \mathrm{e}^{\frac{t''}{\tau^c}}\ \mathrm{d}t'' + \int_0^{n-t'} \mathrm{e}^{-\frac{t''}{\tau^c}}\ \mathrm{d}t''\right) = (\sigma^c)^2 \int_0^n \mathrm{d}t' \left(2 - \mathrm{e}^{-\frac{t'}{\tau^c}} - \mathrm{e}^{-\frac{n-t'}{\tau^c}}\right) =$$
$$= 2(\sigma^c)^2[n + \tau^c(\mathrm{e}^{-\frac{n}{\tau^c}} - 1)] \quad . \tag{3.26}$$

Finally, recalling Eq. 3.23, the following relation is obtained:

$$v^c(n) \equiv n^2 \mathrm{Var}\left[f_p^c(0,n) - G^c g_p^c(0,n)\right] \approx 2(\sigma^c)^2[n + \tau^c(\mathrm{e}^{-\frac{n}{\tau^c}} - 1)] \quad . \tag{3.27}$$

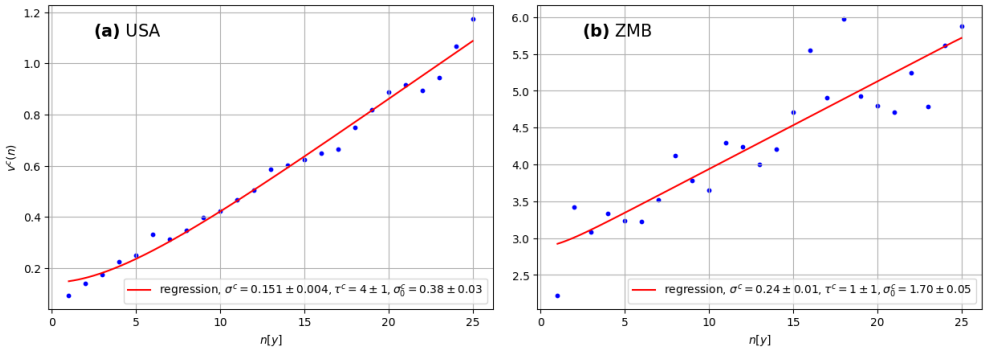

Figure 6: calibration of σ^c, τ^c and σ_0^c in the case of United States (USA, **(a)**) and Zambia (ZMB, **(b)**). The plots provide two examples of the exponential regression of the empirical function $v^c(n)$ (Eq. 3.27) through the exact variance function (Eq. 3.28).

When the analysis is focused on a developed country or on a whole aggregate of countries (like in [4]), fitting the empirical records through the right hand side of Eq. 3.27 usually leads to a reliable estimation of the noise parameters. Conversely, when an underdeveloped country is considered, the distribution of the variances is interested by a relevant positive offset, which cannot enter in the description provided by 3.26 (see Figure 6 for a comparison between the two possible scenarios). Such a displacement is imputable to the poor quality of the approximations involved in the derivation of Eq. 3.27 and describes the influence exerted by the global trade network, considerably strengthened due to the modern globalization process, on the single national economies. In particular, the countries which do not play a predominant role in the international scenario are more susceptible to relevant fluctuations of their trades, which originate from the network of countries they are related to and which manifest themselves as an increase of the aforementioned variances. In order to account for this effect, Eq. 3.27 has been corrected through the addition of the positive shift term $(\sigma_0^c)^2$:

$$v^c(n) \approx 2(\sigma^c)^2[n + \tau^c(\mathrm{e}^{-\frac{n}{\tau^c}} - 1)] + (\sigma_0^c)^2 \equiv w^c(n) \quad . \tag{3.28}$$

As a consequence, the model has been enriched with a new noise parameter, whose calibration provides a more complete insight on the overall fluctuations an export basket can be exposed to.

From the practical point of view, the fit of the empirical records through the exact variance function often leads to the occurrence of convergence issues. Indeed, the optimization of the parameter τ^c is hindered by the extreme related sensitivity of the exponential function and by the resulting presence of a certain number of different minima for the non-linear least squares functions: as a consequence, the choice of the optimal value for τ^c turns out to depend on the initial test values which are set in

the optimization code. However, this inconvenience is easily solved by performing the fit through a proper section of the Maclaurin expansion series of the exact variance function. The implementation of this kind of expansion, which takes the general form

$$w^c_\infty(n) = (\sigma^c_0)^2 + 2(\sigma^c)^2 \sum_{m=1}^{\infty} \frac{(-n)^{m+1}}{(\tau^c)^m (m+1)!} \qquad , \tag{3.29}$$

has the obvious advantage of drastically decreasing the τ^c-sensitivity of the fit. For each country, the fitting code has been executed for increasing orders of the series (starting from the first one) until the achievement of an excellent matching degree between the truncated expansion (calculated on the estimations of the parameters) and the exact function. In particular, the fitting procedure has always been stopped at the accomplishment of the following tuning condition, which regulates the aforementioned matching degree at the furthest end of the domain of the variance function:

$$\left| \frac{w^c_k(25, \sigma^c_k, \tau^c_k, \sigma^c_{0,k})}{w^c(25, \sigma^c_k, \tau^c_k, \sigma^c_{0,k})} - 1 \right| < 0.001 \qquad , \tag{3.30}$$

where σ^c_k, τ^c_k and $\sigma^c_{0,k}$ are the estimations of the parameters at the k-th order. This condition is approached monotonically as the calibration procedure progresses and is still satisfied at orders higher than the final one. More importantly, the final estimations of the parameters proves to be independent of any reasonable initial test values set in the optimization code. In the end, as shown in Figure 6, the imposition of Eq. 3.30 results in a faithful reproduction of the exact variance function and in a reliable estimation of all the parameters.

3.3.3 Calibration of the drift parameter

The calibration of the drift parameter $\overline{\mu}^c$ exploits a particular characterization of the empirical growth defined in Eq. 2.2. After dividing Eq. 3.1 by $X^c_p(t)$, integrating in the interval $[0, T]$ and averaging with respect to both the time and the products, the empirical growth of country c turns out to be decomposed into the following contributions:

$$\lambda^c_T = h^c_T(G^c) + \overline{\mu}^c + \frac{1}{T} \int_0^T I(t)\ \mathrm{d}t \qquad , \tag{3.31}$$

where the term $h^c_T(G^c)$ results from the integration of the coupled part of the model and depends strongly on the transfer parameter. In particular, Eq. 3.31 highlights the direct proportionality between the empirical growth and the drift parameter which exists in the model's perspective. Exploiting this property, the calibration of $\overline{\mu}^c$ can be accomplished, in the first place, by estimating the "reduced" empirical growth,

$$\tilde{\lambda}^c_T \equiv h^c_T(G^c) + \frac{1}{T} \int_0^T I(t)\ \mathrm{d}t \qquad , \tag{3.32}$$

through a significant number of synthetic simulations of the model in absence of the drift contribution ($\overline{\mu}^c = 0$). Then, the subtraction of Eq. 3.32 from Eq. 3.31 provides

the desired estimation:

$$\begin{cases} \lambda_T^c = h_T^c(G^c) + \overline{\mu}^c + \dfrac{1}{T}\displaystyle\int_0^T I(t)\ \mathrm{d}t \\ \tilde{\lambda}_T^c = h_T^c(G^c) + \dfrac{1}{T}\displaystyle\int_0^T I(t)\ \mathrm{d}t \end{cases} \quad \Rightarrow \quad \overline{\mu}^c = \lambda_T^c - \tilde{\lambda}_T^c \qquad , \tag{3.33}$$

where λ_T^c is obtained from the empirical records; the error to be related to the value of the parameter can be estimated indirectly by evaluating the standard deviation of the outcomes of the simulations.

4 Complexity measures of diversity and ubiquity

This chapter provides a detailed technical derivation of the complexity measures of diversity and ubiquity, introduced by Teza, Caraglio and Stella in [14] and extensively analyzed by the same authors in [15]. Through the implementation of entropy, one of the most fundamental and important physical concepts, into the economic complexity field, the measures quantifies in an innovative manner the diversification of the links of the trading bipartite network and allow to deepen the understanding of its intertwined nature. Continuing the branch of analysis started in [4], the complexity measures were originally employed in the aforementioned articles only for the characterization of the export scenario: following this choice and the interpretation of the dynamic model provided in the previous chapter, the following one presents the derivation of the measure from the same perspective. On the other hand, the technical properties of the measures in the import scenario will be briefly exposed in Chapter 6.

The following section describes the definition of the bare versions of the measures and their formal connection with the Shannon's entropy function, providing also a brief report of the line of reasoning which subtends their application; on the other hand, the second section of the chapter supplies a thorough description of the iterative algorithm employed for the refinement of the measures and of its main mathematical properties.

4.1 Bare complexity measures

The key quantity which underlies the entropic characterization of the export bipartite network is the Shannon's entropy function: given an N-dimensional probability vector $\vec{p}$ (such that $\sum_i^N p_i = 1$), the related entropy is defined as

$$H = -\sum_{i=0}^{N} p_i \log\left(p_i\right) \qquad . \tag{4.1}$$

For every fixed $N > 1$, the values assumed by this function are bounded in the range $[0, \log N]$; the lowest one is reached if a probability vector belonging to the canonical basis of $\mathbb{R}^N$ is considered, while the highest one corresponds to a perfectly uniform probability vector (with all its components assuming the value $1/N$). As proved in several recent works related to disparate scientific fields (including ecology [21], economics [22] and abstract mathematics [23]), the range of possible applications of the Shannon's function is very wide and the conceptual meaning of the vectors which define its domain can differ from the original one and can as well be distinguished from the field of abstract probability; still, the soundness of the mathematical properties which characterize the function remains unaltered, regardless of the interpretation of the vectors.

Considering the layer of countries of the bipartite network, the role of probability vector can be played by the set of normalized exports (or *shares*) of a single country in a given year; in order to keep the notation light, the indication of the latter will be omitted for the rest of the chapter. The substitution of the shares of country c, denoted with $\xi_{cp}^{(0)} \equiv X_p^c / \sum_{p'}^{N^c} X_{p'}^c$, into Eq. 4.1 leads to the definition of a basic mea-

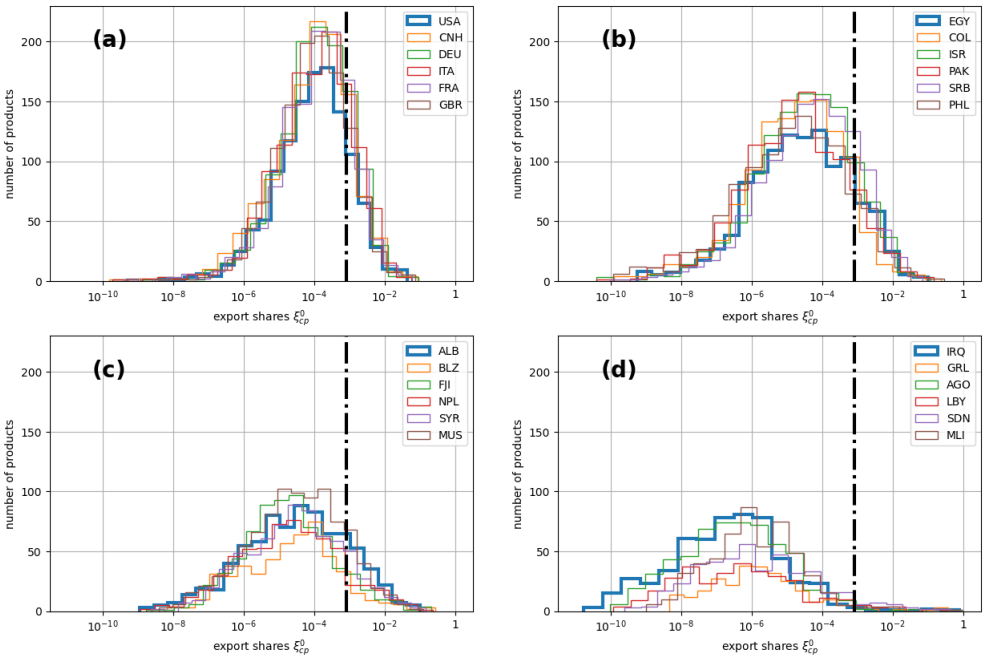

Figure 7: histograms of the distributions of the shares of some countries in year 2020, separated in four economic macro-categories: developed economies **(a)**, semi-developed economies **(b)**, underdeveloped economies **(c)** and risky economies **(d)**. The black dashed lines indicate the ideal limit of a perfectly uniform global distribution.

sure of the *diversity* of its basket:

$$H^{(0)}_{ex,c} = -\sum_{p=0}^{N^c} \xi^{(0)}_{cp} \log\left(\xi^{(0)}_{cp}\right) \quad , \tag{4.2}$$

where N^c denotes the number of products exported. At this level, it is already possible to note that a country with a highly diversified export basket is less exposed to the influence of the global market. Conversely, a country which mainly relies on the production and export of a limited number of products may be particularly vulnerable, for example, to the occurrence of significant fluctuations of the associated prices. It is reasonable to guess that such a stability feature should be typical of the most developed countries, which then should be related to high levels of diversity as defined in Eq. 4.2.

An indication in this sense can be obtained from a preliminary analysis of the distributions of the shares. Figure 7 considers the export baskets of some countries in year 2020 and gathers them in four economic macro-categories, according to a qualitative comparison between the related distributions. Remarkably, the adopted classification

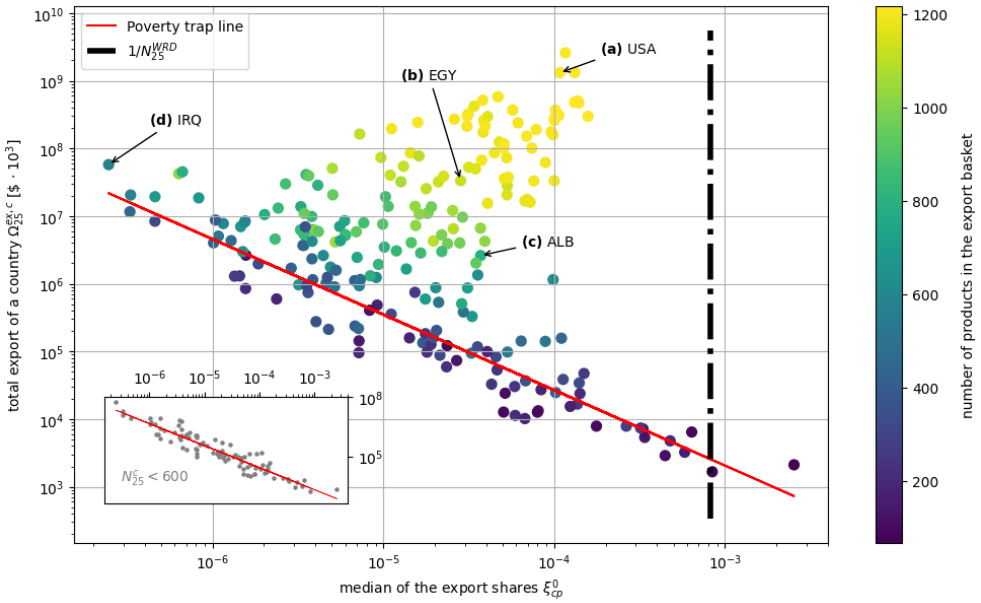

Figure 8: relation between the medians of the distributions of the shares and the total export of each country Ω_{25}^{c} in year 2020. The arrows indicate the position of the representative countries United States (USA, **(a)**), Egypt (EGY, **(b)**), Albany (ALB, **(c)**) and Iraq (IRQ, **(d)**). The red lines in the main plot and in the reduced one highlight the position of the poverty trap area. The black dashed line indicates the ideal limit of a perfectly uniform global distribution.

exhibits an excellent agreement with the macro-economic narrative, as it separates the economies in developed ones (group **(a)**), semi-developed ones (group **(b)**), underdeveloped ones (group **(c)**) and risky ones (group **(d)**). As can be easily noted, the groups differ from each other not only for the average height of the distributions but also for their shapes, which seem to be more distorted as the degree of development decreases. Moreover, the peaks of the distributions in group **(a)** lie in proximity of the ideal value $1/N^{WRD}$, which is peculiar of a perfectly uniform global basket, while in groups **(b)**, **(c)** and **(d)** the peaks are much less defined and more distant from the same value. Figure 8 analyzes the distinction between the shapes of the distributions from a quantitative point of view by comparing the monetary value of all the baskets with the medians of the related distributions. The color scheme applied to the markers reflects the number of products contained in the baskets. The first remarkable property of the plot concerns the disposition of the countries which export less than 600 products, which is limited in a very restricted region characterized by a clear correlation between the quantities represented on the axes. This *poverty trap area* (according to the definition provided in [15]) is related to a condition of poor relevance in the international scenario: not only heavily underdeveloped countries are present (like the group of minor islands in the bottom right region of the plot), but

also countries related to a very valuable basket but limited to a reduced number of similar products (like Iraq with its oil-dependent basket, in the top left region). On the other hand, the plot also suggests, in agreement with Figure 7, that the countries which have managed to detach themselves from the poverty trap area are related to baskets which are generally not only larger both in terms of number of products and monetary value, but also more diversified. For each group of economies represented in Figure 7, the position of a single representative country is indicated with an arrow; the baskets of the four selected countries (United States, Egypt, Albania and Iraq) will be further analyzed in Section 5.3, where the application of the complexity measures will be presented. Considering the positive correlation existing between the median and the Shannon's entropy of a distribution, the above considerations justify the employment of the latter quantity as an estimator of the economic efficiency of the countries, having the additional advantage of incorporating also the information concerning the size of the export baskets.

A similar argument can be supported considering the layer of products. In this case, the probability vector is defined by the set of shares of a given product, $\zeta_{cp}^{(0)} \equiv X_p^c / \sum_{c'}^{N^p} X_p^{c'}$, exported worldwide by the countries. In analogy with Eq. 4.2, the resulting basic measure quantifies the global *ubiquity* of product p and describes the evenness of its distribution in the market:

$$H_{ex,p}^{(0)} = -\sum_{c=0}^{N^p} \zeta_{cp}^{(0)} \log\left(\zeta_{cp}^{(0)}\right) \qquad , \tag{4.3}$$

where N^p is the number of countries participating in the trades.

The measures defined in Eqs. 4.2 and 4.3 represent the bare versions of the final complexity measures which will be employed in Chapter 5 in the analysis of the export bipartite network.

4.2 Refinement iterative algorithm

Despite providing a basic quantification of diversification, the bare measures defined in the previous section do not fully exploit the information supplied by the export database. Both measures are indeed invariant under any permutation of the components of the respective probability vector: in other words, an equal intrinsic weight is assigned to the products contained in the basket of a given country and the same condition holds for the countries which export worldwide a given product. However, in the spirit of the economic complexity approach, countries and products are expected to determine to different extents the values of the corresponding complexity measures. In particular, a very ubiquitous product, exported by the vast majority of the countries, should carry a reduced weight in the evaluation of the diversity of an export basket; analogously, a very diversified national economy, characterized by a comprehensive productive sector, should be less relevant in the quantification of the ubiquity of a product. The intertwined nature of the global trade network can be implemented by means of the construction of an iterative algorithm, having the purpose of progressively refining the complexity measures through a proper re-weighing of the shares; in turn, the adopted re-weighing method should incorporate the information provided

by the complexity measures themselves. The algorithm can be expressed in compact form through the following system:

$$\begin{cases} H_{ex,c}^{(k+1)} = -\sum_{p=0}^{N^c} \xi_{cp}^{(k)} \log\left(\xi_{cp}^{(k)}\right) \\ H_{ex,p}^{(k+1)} = -\sum_{c=0}^{N^p} \zeta_{cp}^{(k)} \log\left(\zeta_{cp}^{(k)}\right) \end{cases} , \tag{4.4}$$

with the shares at the k-th step defined as

$$\xi_{cp}^{(k)} = \frac{X_p^c \; f\left(H_{ex,p}^{(k)}\right)}{\sum_{p'=0}^{N^c} X_{p'}^c f\left(H_{ex,p'}^{(k)}\right)} \quad , \qquad \zeta_{cp}^{(k)} = \frac{X_p^c \; g\left(H_{ex,c}^{(k)}\right)}{\sum_{c'=0}^{N^p} X_p^{c'} g\left(H_{ex,c'}^{(k)}\right)} \quad . \tag{4.5}$$

As shown, the progressive re-weighing of the shares is accounted for by the introduction of the functions f and g. In order to illustrate in a clear manner, for example, the role played by g on the refinement of the ubiquity, it is convenient to consider the generic monopolistic scenario consisting in the exclusive export of product p by country c ($\zeta_{cp}^{(0)} \approx 1$, while $\zeta_{c'p}^{(0)} \approx 0 \; \forall c' \neq c$). In order to keep the notation light, the images through f and g of the bare complexity measures will be hereafter considered as independent variables and denoted accordingly: $f\left(H_{ex,p}^{(0)}\right) \to f_p$ and $g\left(H_{ex,c}^{(0)}\right) \to g_c$. Looking at the structure of the shares in Eqs. 4.5 and considering the behavior of the function $-x\log(x)$ in the interval $[0,1]$, it is easy to observe that an increase (respectively, decrease) of g_c leads to a coherent decrease (increase) of $H_{ex,p}^{(1)}$, due both to the increasing (decreasing) relevance of the related share ($\zeta_{cp}^{(0)} \to 1$) and the increasing (decreasing) irrelevance of the other ones ($\zeta_{c'p}^{(0)} \to 0 \; \forall c' \neq c$). Obviously, an analogous argument holds in the case of an export basket dominated by a single product and the resulting effects are in both cases amplified if more steps of the algorithm are performed. Although the influence of more general changes in the sets $\{f_p\}$ and $\{g_c\}$ on the complexity measures in not always fully coherent and is then less obvious, the same conclusions are still valid if more regular distributions of countries and products are considered. Based on these considerations, in order to implement a proper re-weighing of the shares, f and g should be defined as monotonically decreasing functions of the respective complexity measures. The choice of the formal dependence is not unique and should reflect the aim of the analysis. Following the prescription adopted in [15], which exploits the boundedness of the Shannon's entropy function, the re-weighing functions are here defined as simple linear relations:

$$f\left(H_{ex,p}^{(k)}\right) = \log\left(N^c\right) - H_{ex,p}^{(k)} \quad , \qquad g\left(H_{ex,c}^{(k)}\right) = \log\left(N^p\right) - H_{ex,c}^{(k)} \quad . \tag{4.6}$$

In particular, the most relevant property which follows from the linearity and the strict positivity of these functions is the global continuity of the shares defined in Eq.

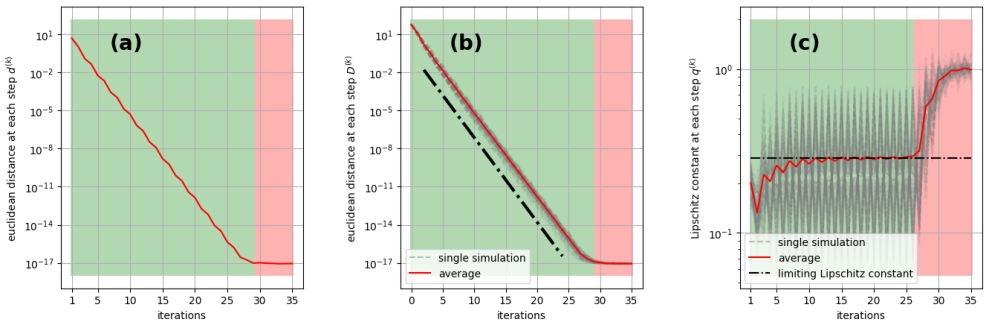

Figure 9: convergence properties of the iterative algorithm in the export scenario: Euclidean distance between consecutive steps $d^{(k)}$ (Eq. 4.8) **(a)**, Euclidean distance from the empirical fixed point $D^{(k)}$ (Eq. 4.9) **(b)** and Lipschitz constant $q^{(k)}$ related to the map ϕ (Eq. 4.7) **(c)**. In panels **(b)** and **(c)**, the red lines highlight the evolution of the mean Euclidean distances extracted from a set of 100 simulations.

4.5. As a consequence, the algorithm can be mathematically characterized as a continuous map ϕ of a subset S onto itself, regardless of the number of steps implemented:

$$\phi: \; S := \prod_{i=0}^{C} [0, \log(N^{c_i})] \times \prod_{j=0}^{P} [0, \log(N^{p_j})] \to S \tag{4.7}$$

where C and P are the numbers of countries and products defining the export bipartite network and each Π symbol expresses in compact form a series of Cartesian products. Moreover, the compactness and the convexity of S, which follow from its characterization as a bounded and closed hyper-parallelepiped embedded in $\mathbb{R}^C \times \mathbb{R}^P$, allow to state the existence of at least one fixed point for ϕ in $S \backslash \partial S$ as a result of the application of the celebrated Brouwer's theorem [39]. In turn, proving the attractive nature and the uniqueness of such a point allows to demonstrate the global convergence and the stability of the algorithm. The first property has been proved by studying the evolution of the Euclidean distance between the complexity measures related to consecutive steps of the algorithm, defined as

$$d^{(k)} = \sqrt{\sum_{c=0}^{C} \left(H_{ex,c}^{(k+1)} - H_{ex,c}^{(k)} \right)^2 + \sum_{p=0}^{P} \left(H_{ex,p}^{(k+1)} - H_{ex,p}^{(k)} \right)^2} \tag{4.8}$$

for the k-th step. Figure 9a depicts the behavior of this distance when the empirical bare complexity measures (Eqs. 4.2 and 4.3) are set as initial conditions of the algorithm and allows to observe their exponentially fast approach towards a point belonging to the inner region of S.

The next step consists in checking if S is fully contained in the basin of attraction of this point (hereafter denoted with $(H_{ex,c}, H_{ex,p})$), a property that would exclude the presence of other attractive fixed points for ϕ. To this purpose, the algorithm has been applied to a sample of 100 points randomly selected from the inner region of S;

then, the Euclidean distance between the complexity measures at a given step and $(H_{ex,c}, H_{ex,p})$, defined as

$$D^{(k)} = \sqrt{\sum_{c=0}^{C} \left(H_{ex,c}^{(k)} - H_{ex,c} \right)^2 + \sum_{p=0}^{P} \left(H_{ex,p}^{(k)} - H_{ex,p} \right)^2} \tag{4.9}$$

for the k-th step, has been evaluated and found to be, on average, exponentially converging to 0 (Figure 9b). Remarkably, the monotonic nature of the convergence of each single synthetic distance indicates the absence of any repulsive fixed point for ϕ in S. Lastly, the evaluation of the asymptotic Lipschitz constant related to ϕ, estimated as

$$q = \lim_{k \to \infty} \frac{D^{(k+1)}}{D^{(k)}} \quad , \tag{4.10}$$

has allowed to describe the convergence of the complexity measures from a different perspective. The calculation has been performed on the same sample of 100 points already employed and the outcome of each simulation, along with the mean one, is represented in Figure 9c; the value of the asymptotic constant, $q_{ex} \approx 0.28$, has been found to be in agreement with the characterization of ϕ as a contraction. The above considerations allow to state the convergence of the iterative algorithm towards the unique fixed point $(H_{ex,c}, H_{ex,p})$ and the validity of the following consistency relations:

$$\begin{cases} H_{ex,c} = -\sum_{p=0}^{N^c} \xi_{cp} \log\left(\xi_{cp}\right) \\ H_{ex,p} = -\sum_{c=0}^{N^p} \zeta_{cp} \log\left(\zeta_{cp}\right) \end{cases} \quad , \tag{4.11}$$

with the re-weighted shares

$$\xi_{cp} = \frac{X_p^c \; f\left(H_{ex,p}\right)}{\sum_{p'=0}^{N^c} X_{p'}^c f\left(H_{ex,p'}\right)} \quad , \qquad \zeta_{cp} = \frac{X_p^c \; g\left(H_{ex,c}\right)}{\sum_{c'=0}^{N^p} X_p^{c'} g\left(H_{ex,c'}\right)} \quad . \tag{4.12}$$

Ultimately, the application of the iterative algorithm provides an estimation of the intrinsic relevance of the nodes which constitute the export bipartite network. Adopting a different perspective, the progressive iteration of Eq. 4.4 may also be interpreted as an attempt to gradually determine the weight of the factors which determine the global productive system. From this point of view, the diversity of an export basket offers information concerning the quality of the productive infrastructure which supports it, while the ubiquity of a product acts as an inverted indicator of the technological sophistication required for its development. Usually, the purpose of obtaining a reliable assessment of the steps which lead to the realization of a single product is quite prohibitive. The major obstacle encountered in this task is the high degree of

external contamination which affects, as a consequence of the modern globalization process, the single productive chains, whose intermediate steps are often spread across a multitude of different countries. The approach presented in [15] and reported in this chapter offers a natural solution to this issue by considering from the beginning the intertwined nature of the export bipartite network and, indirectly, the ensemble of interconnections typical of the global productive system.

5 Analysis of the export bipartite network

Following the narrative of the articles by Caraglio, Baldovin, Stella and Teza [4,14,15], this chapter supplies an exhaustive presentation of the main results deriving from the application in the export scenario of the theoretical models introduced previously. The main subject of the first section is the validation of the dynamic model through the demonstration of its ability to reproduce some of the most important features which characterize the export database; then, the subsequent section presents an application of the model consisting in a counterfactual analysis of the growth prospects of the countries, leading to an innovative interpretation of the role played by the transfer of resources between different market sectors. Afterwards, the chapter changes perspective and dedicates a whole section to the characterization of the export bipartite network by means of the application of the complexity measures, delving into the main properties of both its layers. Lastly, the final section of the chapter presents a novel dynamical characterization of the network of countries resulting from the unification of the dynamical and the entropic branches of analysis.

5.1 Validation of the dynamic model

Given the incorporation of some features of the export database into the structure of the dynamic model (like the formation of a specific ranking of the products over time and the cross-correlation between the evolution of the exports), the results presented in this paragraph serve primarily as consistency checks.

In the first place, Figure 10 depicts the synthetic reproduction, at the 2-digit aggregation level, of the export baskets from United States, China and Russia. Each plot presents a single simulation of the model, performed with the implementation of the calibrated parameters. As can be easily noted from a quick comparison with Figure 2, each basket is globally interested by an exponential growth which faithfully recalls

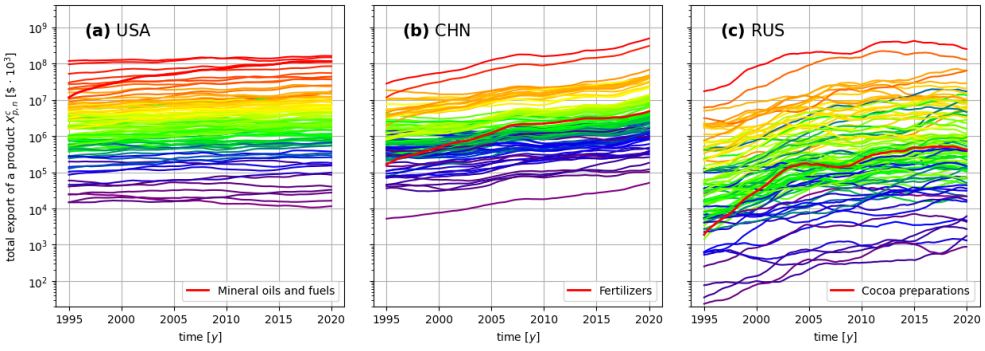

Figure 10: synthetic evolution across the period 1995-2020 of the export baskets of United States (USA, **(a)**), China (CHN, **(b)**) and Russia (RUS, **(c)**) at the 2-digit aggregation level. The wavelength of the color assigned to each product reflects its rank (Eq. 2.1) inside the respective basket. For each country, the product interested by the most relevant growth (in relative terms) is marked with a thick red line.

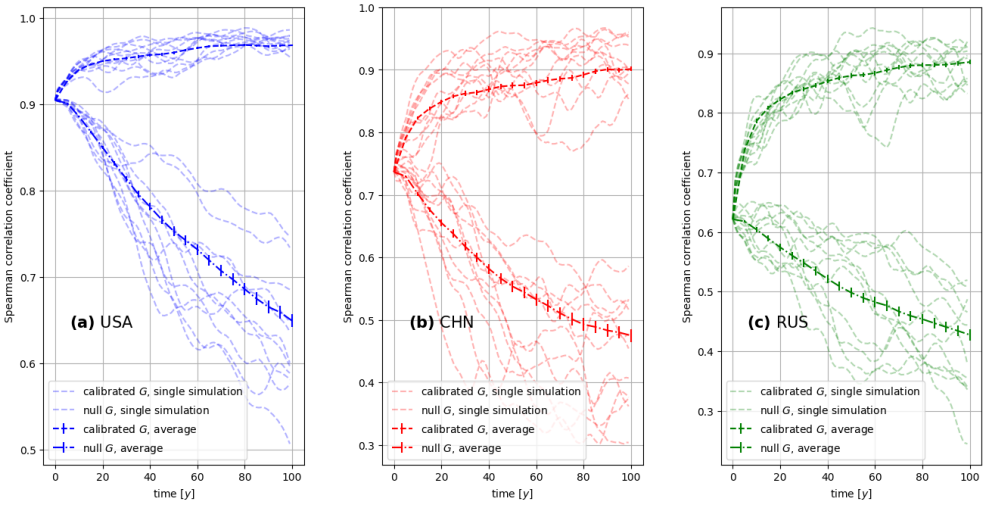

Figure 11: evolution of the Spearman's correlation coefficients ρ_S (Eq. 5.1) across a 100-years period starting from 1995 related to the rankings of United States (USA, **(a)**), China (CHN, **(b)**) and Russia (RUS, **(c)**. In each plot, the thick dashed lines describe the mean evolution of the coefficient extracted from two distinct sets of 100 simulations, respectively associated to the calibrated values of G^c and to null values of the same parameters.

its respective empirical counterpart, with China's one considerably stronger than the other two; additionally, the synthetic volatility also exhibits a close similarity with the empirical one. Like in Figure 2, the wavelength of the color assigned to each product reflects its rank inside the respective empirical ranking (Eq. 2.1): the implementation of this feature generates a strengthening rainbow effect clearly perceivable in every plot, confirming the ability of the model to progressively reproduce the correct final distribution of the exports. In this regard, the plots also highlight the evolution of the same products marked in Figure 2, which, in agreement with the historical trend, are distinguished among the others for an above-average relative growth rate.

A way to quantitatively assess the quality of reproduction of the ranking consists in the evaluation of the Spearman's rank correlation coefficient, denoted with ρ_S. The purpose of this coefficient is the assessment of the likelihood of a monotonic relation potentially existing between two variables and can thus be interpreted as a generalization of the usual Pearson's correlation coefficient. As its name suggests, this task is accomplished by considering the ranks of the values assumed by the variables in place of the values themselves: if all the ranks, denoted with rk, appear as distinct integers, the Spearman's coefficient of two random variables X and Y can be calculated by means of the simplified formula

$$\rho_S = 1 - \frac{6}{N(N^2-1)} \sum_{n=0}^{N} (\mathrm{rk}(x_n) - \mathrm{rk}(y_n))^2 \quad , \tag{5.1}$$

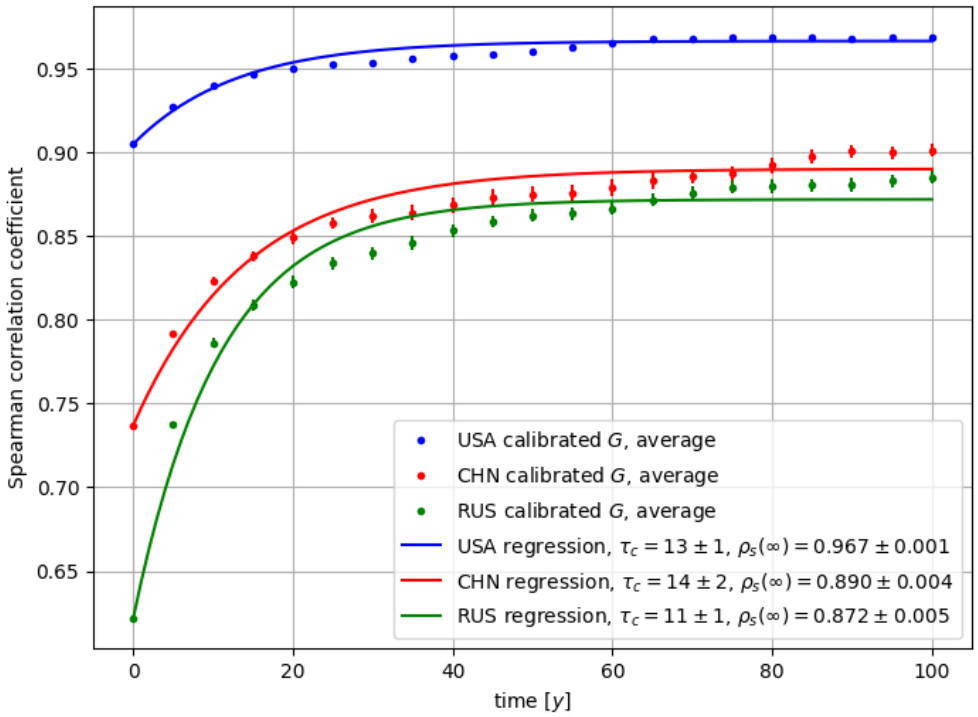

Figure 12: calibration of τ_c and $\rho_S^c(\infty)$ in the case of United States (USA), China (CHN) and Russia (RUS).

where N is the number of total observations. The value of ρ_S is always bounded between -1 (exact negative correlation between the rankings) and $+1$ (exact positive correlation); obviously, a null value indicates that the rankings are completely uncorrelated. In the context of the validation of the dynamic model, the interest lies in the evolution over time of the Spearman's coefficient between the synthetic ranking (again defined as an average across a five-years period) and the empirical one. In order to apply Eq. 5.1 it has been first necessary to convert the fractional ranks provided by the shares into integer ones: since it is extremely unlikely to observe identical ranks within any of the two rankings, this task has been accomplished by simply matching every product to the corresponding placement (for example, 1 is assigned to the most exported product). Then, a set of 100 simulations of the dynamic model has been performed across a 100-years period in the case of United States, China and Russia; the evolution of the resulting mean values of ρ_S is depicted in Figure 11, along with the outcomes of reduced samples of simulations. If the calibrated value of G^c is set, the convergence of the synthetic ranking towards the empirical one is immediately recognizable in every plot and represents a confirmation of the soundness of the particular form adopted for the transfer matrix (Eq. 3.16). However, it is also easy to note that the asymptotic values of the mean Spearman's coefficients provided by the

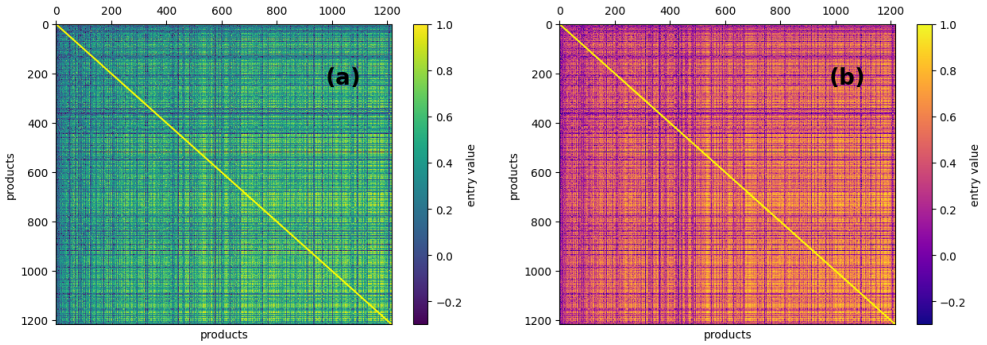

Figure 13: comparison between the empirical **(a)** and the synthetic **(b)** correlation matrices related to the exports of the countries aggregated at the global level. In both cases, the products are distributed according to their global ranks.

model never coincide with the unit one and that they are not even reached within the 25-years period covered by the database; the responsibility for both these properties is attributable to the random fluctuations, which are particularly relevant when large baskets are considered and are also at the origin of the great variance of the simulations observed in Figure 11. Furthermore, given the exponentially suppressed growth which interests each of the mean values of the Spearman's coefficient, a characteristic time τ_c can be related to each country by performing a fit with the exponential function

$$\rho_S^c(t) = \rho_S^c(\infty) + [\rho_S^c(0) - \rho_S^c(\infty)]\mathrm{e}^{-\frac{t}{\tau_S^c}} \qquad , \tag{5.2}$$

which also provides an estimation of the asymptotic values provided by the model. For each country, the outcomes of the regressions are shown in Figure 12. As an indicator of the rapidity of the synthetic ranking's variation, the characteristic time in Eq. 5.2 can be interpreted, from a macro-economic perspective, as a measure of the resistance opposed by the economy of the country to particular substantial alterations of its structure.

Figure 11 also depicts the behavior of the Spearman's coefficient if a null value of G is set. As expected, in each plot the related set of curves follows a completely different trend and coherently evolves towards a condition of complete uncorrelation between the ranking involved. This last feature, in particular, allows to appreciate the importance of the implementation of a coupled section in the model 3.1, in order to properly replicate the historical evolution of a given basket.

Changing perspective, another confirmation of the consistency of the dynamic model comes from the synthetic reproduction of the empirical time correlation between the exports. In this regard, Figure 13 realizes a comparison between the empirical correlation matrix (already shown in Figure 3) and a synthetic counterpart extracted from a set of 100 simulations, distributing in both cases the products according to their global ranks. Again, the immediately notable high degree of symmetry between panels *(a)* and *(b)* allows to confirm the ability of the dynamic model to reproduce all

the empirical correlation structures which characterize the export database and proves the validity of the correlation properties incorporated into the structure of the model itself.

Lastly, in order to gain a deeper insight on the role played by the parameters of the model, it is interesting to amplify or to suppress their influence and to display the consequent alterations on the evolution of a given export basket. In this regard, Figure 14, Figure 15 and Figure 16 consider the example of the basket of United States at the 2-digit aggregation level. First of all, from a quick comparison between Figure 14 and Figure 10a, it is easy to note the extreme effects of the attraction force carried out by the empirical ranking on the whole basket if an anomalously high value of G^c is set: as expected, the excessive relevance of the coupled part of the dynamic model is

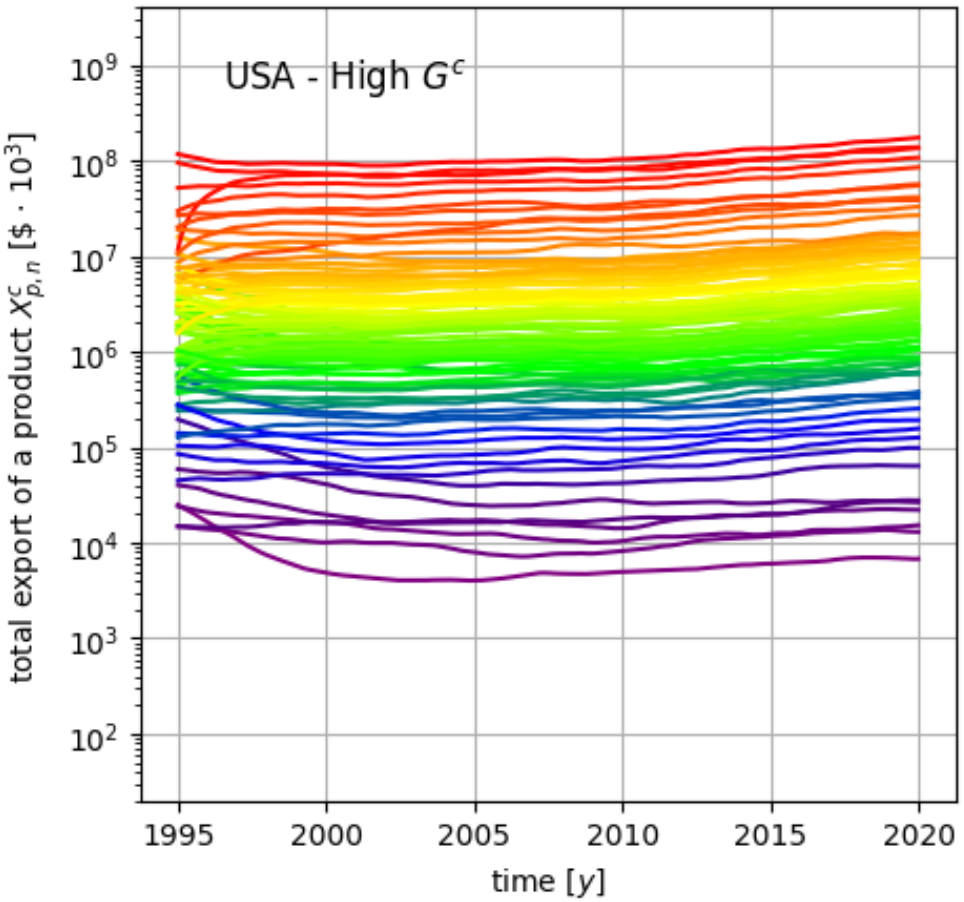

Figure 14: synthetic evolution across the period 1995-2020 of the export basket of United States (USA) resulting from the implementation of a 10 times higher value of G^c.

responsible for a very premature redistribution of the resources. Then, turning to the properties of the noise term, the subsequent plots in Figure 15 unveil the influence of σ^c and τ^c on the volatility of the evolution. In agreement with Eq. 3.7, an amplified value of σ^c is related to an increased time correlation in the evolution of the noise term and leads to the presence of long, coherent periods of persistent growth or recession; on the other hand, setting a depressed value of τ^c makes any correlation interval shrink and causes the production of extremely noisy plots. Conversely, in agreement with

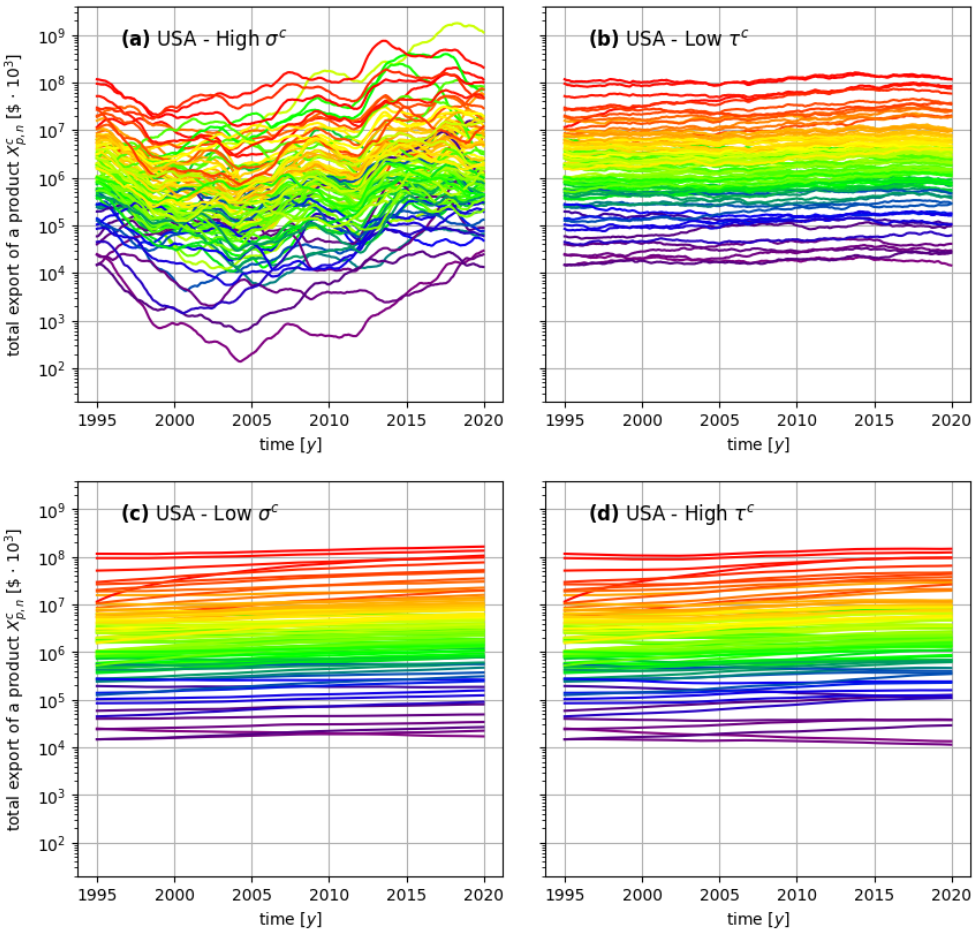

Figure 15: synthetic evolution across the period 1995-2020 of the export basket of United States (USA) resulting from the implementation of a 10 times higher value of σ^c **(a)**, a 10 times lower value of τ^c **(b)**, a 10 times lower value of σ^c **(c)** and a 10 times higher value of τ^c **(d)**.

the equations which regulate the evolution of the noise term (Eqs. 3.13 and 3.14), taking σ^c or τ^c to the opposite extremes completely nullifies any form of volatility and produces totally flat reproductions. Finally, the last plot in Figure 16 shows the

expected increase of the exponential growth of the whole basket which directly results from the amplification of $\overline{\mu}^c$.

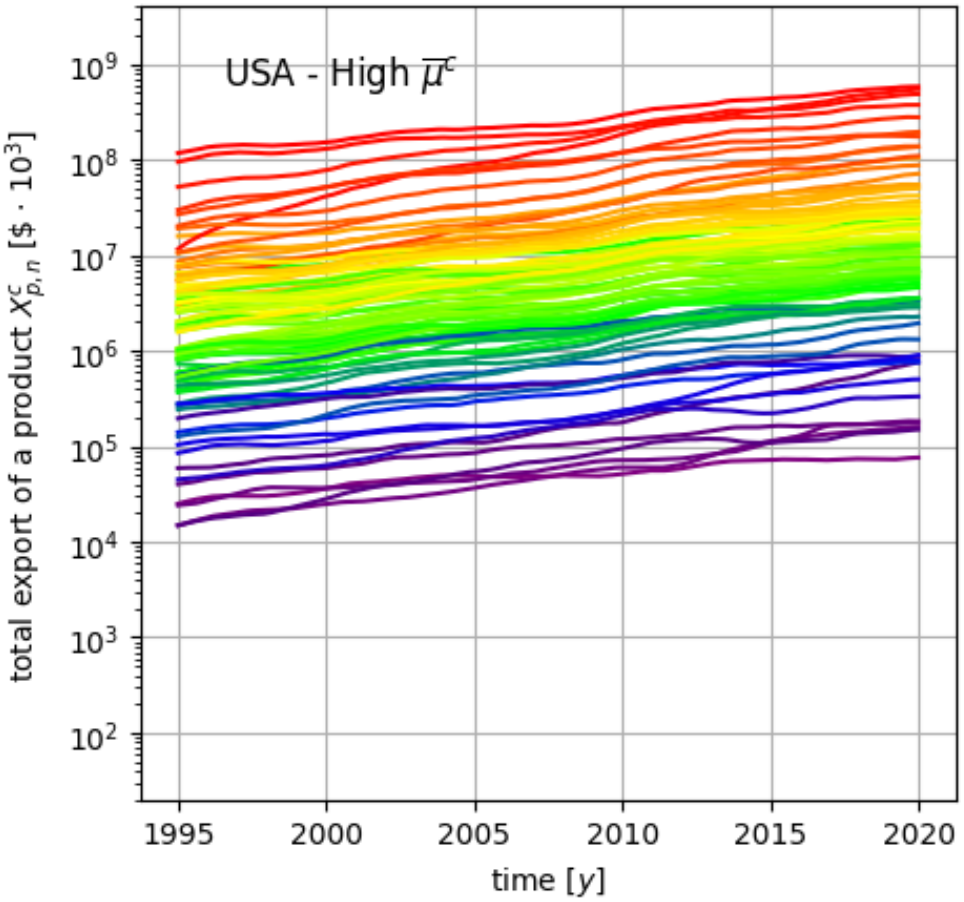

Figure 16: synthetic evolution across the period 1995-2020 of the export basket of United States (USA) resulting from the implementation of a 10 times higher value of μ^c.

5.2 Counterfactual analysis of growth prospects

At the end of Section 3.3, the procedure of calibration of the drift parameter $\overline{\mu}^c$ has exploited the following particular decomposition of the empirical growth λ^c_T (Eq. 2.2):

$$\lambda^c_T = h^c_T(G^c) + \overline{\mu}^c + \frac{1}{T}\int_0^T I(t)\ \mathrm{d}t \qquad . \tag{5.3}$$

This relation highlights the multifactorial nature which characterizes λ^c_T from the per-

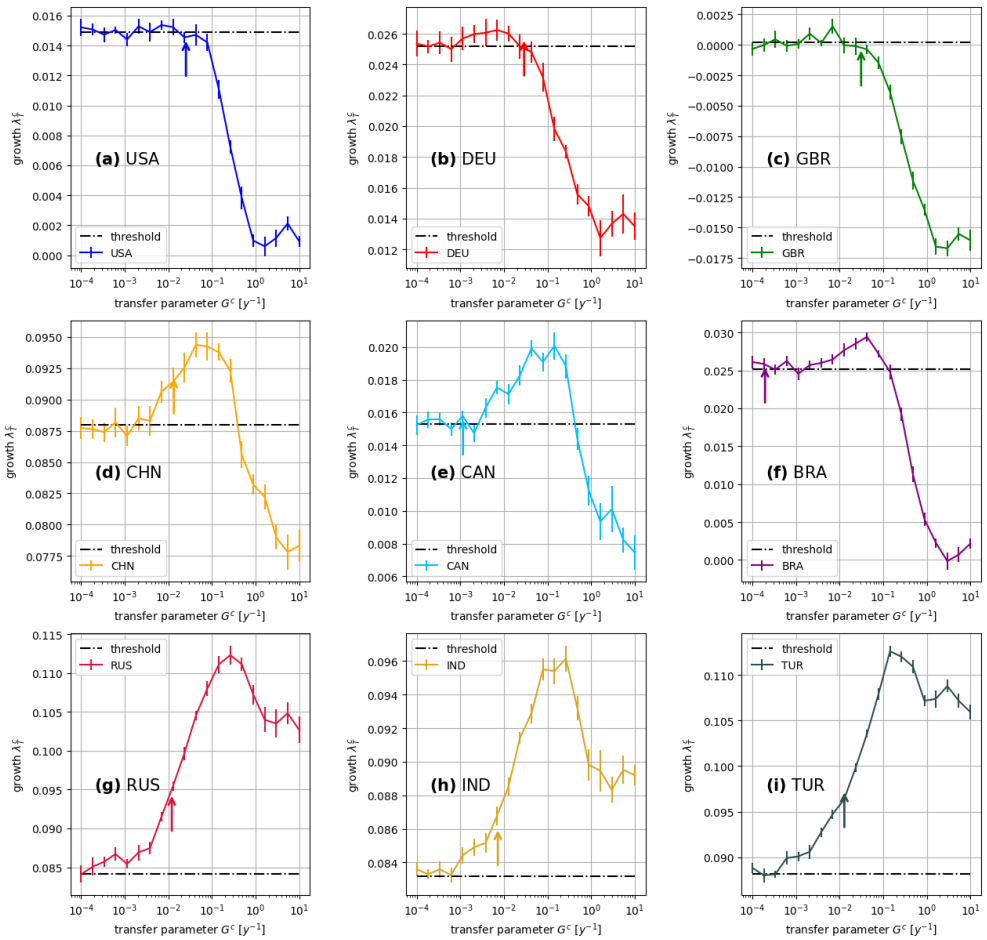

Figure 17: counterfactual behavior of the growth λ_T^c in the case of United States (USA, **(a)**), Germany (DEU, **(b)**), United Kingdom (GBR, **(c)**), China (CHN, **(d)**), Canada (CAN, **(e)**), Brazil (BRA, **(f)**), Russia (RUS, **(g)**), India (IND, **(h)**) and Turkey (TUR, **(i)**). In each plot, an arrow indicates the calibrated value of G^c and the empirical value of λ_T^c, while the black dashed line marks the threshold of the growth related to a null contribution from the transfer mechanism.

spective of the dynamic model. Interestingly, the growth of an export basket is not entirely attributable to the constant drift term and to the natural increase of the prices over time, but an additional significant contribution originates from the presence of the transfer mechanism incorporated into the model. This peculiar effect stems from the occurrence of local favorable stochastic fluctuations, which then propagate across the whole network of products outclassing the unfavorable ones, due the multiplicative nature of the terms which govern the equations. The actual extent of the fluxes of resources between different exports is controlled by the calibrated value of transfer

parameter G^c, which results from the procedure described in Section 3.3. However, an additional interesting insight concerning the economy of a country can be gained by studying the behavior of the overall growth in response to an alteration of G^c alone, leaving the remaining parameters untouched. The interest in this branch of analysis mainly stems from the prescriptive nature of G^c, which can be in principle partly controlled through the adoption of particular regulations and investment policies. Moreover, the tuning of the transfer mechanism represents an example of exploration-exploitation trade-off problem [11–13], defined by the choice between fully exploiting a positive trend related to a single product or exploring the possible usefulness of other similar trends by shifting some resources from its production: in this perspective, the branch of analysis presented here serves as a retrospective counterfactual study of the overall growth which could have interested the export baskets if more flexible internal investment policies had been adopted throughout the evolution.

The plots in Figure 17 depict the behavior of the growth of the baskets of some relevant countries, which are gathered according to the three main qualitative scenarios that can be encountered. In each plot, an arrow indicates the calibrated value of G^c and the empirical value of λ_T^c, which is always found to be in excellent agreement with the curves; furthermore, the black dashed line marks the threshold of the growth related to a null contribution from the transfer mechanism and helps to assess the relevance of the latter as G^c is increased. Every growth value results from a set of 1000 simulations. The most relevant feature which distinguishes the three categories of plots is the relation between the growth curves and the threshold lines. In the case of the USA-like group (panels **(a)**, **(b)** and **(c)**), the action of the transfer mechanism barely manages to provide a positive contribution to the overall growth in the intermediate region of values and becomes a severe hinder when the value of G^c overcomes the calibrated one. On the other hand, the remaining plots depict a scenario where the aforementioned mechanism is able to significantly enhance the overall growth up to a peak which always lies beyond the calibrated value of G^c; while in the China-like plots (panels **(d)**, **(e)** and **(f)**) an excessive transfer of resources can make the growth decline below the threshold, in the Russia-like one (panels **(g)**, **(h)** and **(i)**) such a phenomenon is always able to provide a positive contribution. These differences are ascribable to the degree of stability which characterizes the economy of a country during the first years of the period of interest. The first row of plots represents national economies presenting optimal, consolidated structures which were able to prosper by fully exploiting any local positive trend; the second row, instead, represents developing economies which would have benefit from a moderately higher degree of cooperation between similar productive chains; finally, the third row represents underdeveloped economies or economies subjected to drastic structural changes at the end of the last century (with the emblematic case of Russia, after the crumbling of the communist bloc), which would have gained a great advantage from an active mechanism of exploration of correlated positive trends.

Lastly, in order to fully appreciate the relative contribution provided by the transfer mechanism, Figure 18 presents all the curves deprived of both the constant drift term and the inflationary contribution, assigning a different color to each group of countries.

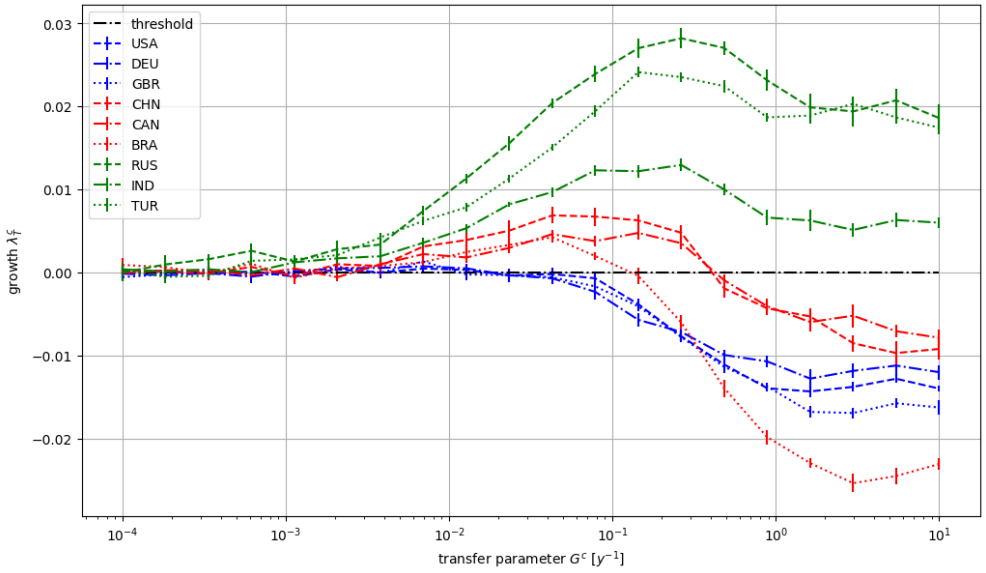

Figure 18: counterfactual behavior of the growth λ_T^c deprived of the constant drift term and the inflationary contribution in the case of United States (USA), Germany (DEU), United Kingdom (GBR), China (CHN), Canada (CAN), Brazil (BRA), Russia (RUS), India (IND) and Turkey (TUR).

5.3 Diversification of the bipartite network

This section presents the most interesting results emerging from the application of the final complexity measures in the export scenario. The representation of the fixed point reached by the iterative algorithm defined in Section 4.1 will be followed by the analysis of the connection between the measures related to different aggregation levels, in order to gain an insight on the inner degree of organization of each national economy.

5.3.1 Refined complexity measures

The plots in Figure 19 provide a comparison between the final complexity measures and the total export related to the nodes of the export bipartite network in year 2020, separating the layer of countries from the one of the products.

In Figure 19a, the color applied to each marker reflects the number of products exported by the respective country. In light of the discussion concerning the distributions of the shares in Section 4.2 and observing Figure 8, it's not surprising that the countries which play a major role in the global trade network are characterized by a larger and more diversified export basket. On the opposite side, the countries with baskets containing less than 600 products, confined to the poverty trap area of Figure 8, are spread over the rather large area highlighted in red and assume consistently lower diversity values. In addition, the plot indicates the position of the

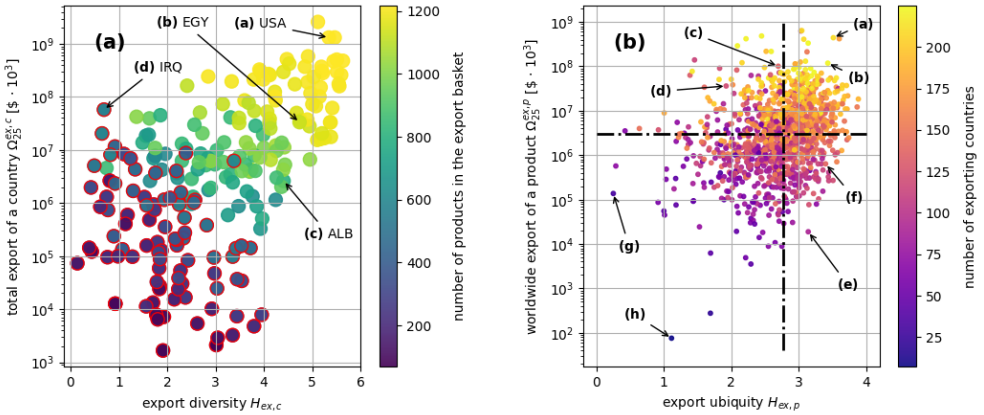

Figure 19: relation between the final complexity measures and the total export of each node of the export bipartite network, in year 2020. In panel **(a)**, the markers circled in red describe the region occupied by the poverty trap area, while the arrows indicate the position of the four representative countries United States (USA), Egypt (EGY), Albania (ALB) and Iraq (IRQ). In panel **(b)**, the black dashed lines refer to the medians of the quantities represented on the axes and subdivide the global market in four regions, while the arrows indicate the position of non crude oils **(a)**, insulated electric conductors **(b)**, aircrafts and spacecrafts **(c)**, photographic laboratory equipment **(d)**, hand-woven tapestries **(e)**, paints and varnishes **(f)**, cobalt ores **(g)** and phosphides **(h)**.

same four countries (United States, Egypt, Albania and Iraq) previously chosen as representatives of each macro-category of national economies (Figure 7). As can be easily noted, United States and Egypt occupy a position in the corner of most developed countries and present similar degrees of diversity of the respective baskets; on the other hand, Albania and Iraq are clearly detached from the top right corner and assume, respectively, a slightly lower rank and one of the last ranks inside the diversity ranking. Notably, despite having a significantly less valuable basket, Albania is clearly distinguished from Iraq due to the more regular distribution of its exports, which allows a clear detachment from the poverty trap area. These considerations express the ability of the diversity measure to highlight in a straightforward manner both the difference between developed and under-developed countries and the distinction between the different kinds of economies and therefore allow to consider it as a better efficiency estimator than the median of distributions of the shares (according to which, for example, Egypt and Albania are placed at an equal distance from the ideal limit of a perfectly uniform global distribution).

On the other hand, Figure 19b provides a representation of the layer of products by means of the ubiquity measure. Similarly to the previous case, each marker is colored according to how many countries export the respective product. The black dashed lines describe the medians of the distributions of the quantities represented on the axes and allow to subdivide the global market into four main regions. The top right one

contains products with large volume of global export and high ubiquity. Due to their relevance in the market, the products which fall in this section (like non crude oils and insulated electronic machinery) represent the cornerstones of the global economy and are generally exported by the vast majority of the countries. On the other hand, the products in the top left region are characterized by an equally large amount of export but assume significantly lower ubiquity values. When exported by a limited number of countries, some of these products (like spacecrafts and photographic laboratory equipment) are representatives of very specialized and technologically sophisticated productive chains and indicate the presence of relevant oligopolistic market sectors. Finally, the lower side of the plot collects the less exported products, whose relevance in the market decreases as the region of low ubiquity values is approached.

At this point, it is interesting to analyze the overlap degree between the global export basket represented in Figure 19b and the baskets of some countries of interest. In this regard, Figure 20 realizes such comparison in the case of the four representative countries United States, Egypt, Albania and Iraq. In each of the four plots, the products which belong to the respective national basket are marked with a specific color, while the external products are marked in black. The four percentages reported in each plot provide a measure of the overlap between the two baskets in each of the four main regions of the market. Also, every plot indicates with dashed lines the medians of the distributions of both the total export and the ubiquity of the products contained in the given national basket, in order to compare them with the global ones. The first feature that can be noted observing Figure 20 is the extreme distance in terms of the number of exported products which separates the four countries considered: while United States exports every product in the market and Egypt has a very comprehensive basket, the baskets from Albania and Iraq present a very poor overlap with the global one. Furthermore, as the size of a basket decreases, it is possible to appreciate the progressive shift of the related medians towards the region of most exported and ubiquitous products: notably, in the case of Albania and Iraq, the only significant percentage is related to the most relevant market region, while the occupation of the other ones is moderately or even extremely scarce. This feature clearly shows how less developed countries seem to concentrate their productive efforts on the most dominant products in the market: this fact is not surprising, since the lack of a properly sophisticated productive infrastructure obviously forces such countries to invest in products which are relatively easy to produce and also highly demanded in the market (both of these properties are indeed at the origin of the high ubiquity of such products). On the other hand, it is quite intriguing to note the tendency by the most developed countries to export a number as large as possible of different products, including the most marginal ones. This peculiar feature highlights in a deeper way the role played by the diversity deducible from Figure 19a, realizing a clearer distinction between the different macro-categories of national economies. As a remarkable conclusion, in order to reach or maintain optimal degrees of development and international competitiveness, a country can hardly refrain to keep its export basket as diversified as possible by concentrating a part of its productive efforts also on poorly demanded products; conversely, a lack of the latter in the export basket of a country can be safely considered as a symptom of underdevelopment and general irrelevance in the international scenario.

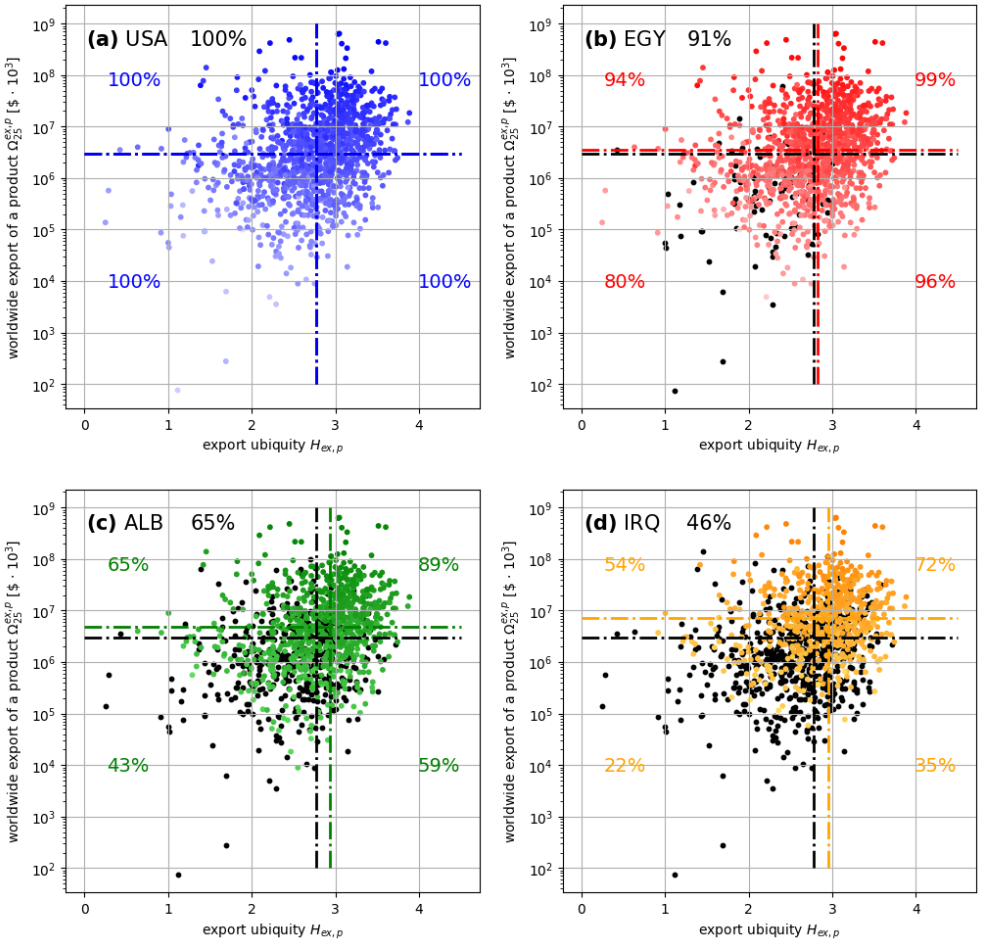

Figure 20: overlap between the global basket and the baskets from United States (USA, **(a)**), Egypt (EGY, **(b)**), Albania (ALB, **(c)**) and Iraq (IRQ, **(d)**), in year 2020. The colored dashed lines indicate the medians of the distributions of the quantities on the axes associated to the products contained in the national baskets. For each global market region, a percentage indicates the degree of overlap between the national baskets and the global one.

5.3.2 Intra-sectorial and inter-sectorial diversity contributions

The peculiar structure of the Shannon's entropy function allows to build simple relations between the diversities associated to different levels of aggregations. So far, both the dynamics and the entropic branches of analysis have concerned the intermediate aggregation level of BACI, which collects in a total of approximately 1200 categories (depending on the year considered) the products which share the four most significant classification digits. Turning the focus on the 96 macro-categories of products which

constitute the 2-digits aggregation level, the final shares defined in Eq. 4.12 can be implemented into the definition of the following coarse-grained inter-sectorial diversity measure:

$$H_{ex,c}^{CG} = -\sum_{P} \xi_{cP} \log\left(\xi_{cP}\right) \quad , \tag{5.4}$$

where each of the quantities $\xi_{cP} \equiv \sum_{p \in P} \xi_{cp}$ defines the total share of a single macro-category. On the other hand, the inner degree of diversification of the generic macro-category P can be investigated through the definition of an intra-sectorial diversity measure:

$$H_{ex,c}^{P} = -\sum_{p \in P} \frac{\xi_{cp}}{\xi_{cP}} \log\left(\frac{\xi_{cp}}{\xi_{cP}}\right) \quad , \tag{5.5}$$

which is obviously consistent with the normalized nature of the shares. Then, a quick calculation can easily show that the difference between the fine diversity $H_{ex,c}$ and the coarse one $H_{ex,c}^{CG}$ can be expressed as a weighted average of all the intra-sectorial contributions:

$$\begin{aligned}
\Delta H_{ex,c} \equiv H_{ex,c} - H_{ex,c}^{CG} &= -\sum_{p} \xi_{cp} \log\left(\xi_{cp}\right) + \sum_{P} \xi_{cP} \log\left(\xi_{cP}\right) = \\
&= \sum_{P} \left(-\sum_{p \in P} \xi_{cp} \log\left(\xi_{cp}\right) + \xi_{cP} \log\left(\xi_{cP}\right) \right) = \\
&= \sum_{P} \xi_{cP} \left(-\sum_{p \in P} \frac{\xi_{cp}}{\xi_{cP}} \log\left(\xi_{cp}\right) + \log\left(\xi_{cP}\right) \sum_{p \in P} \frac{\xi_{cp}}{\xi_{cP}} \right) = \\
&= -\sum_{P} \xi_{cP} \sum_{p \in P} \frac{\xi_{cp}}{\xi_{cP}} \log\left(\frac{\xi_{cp}}{\xi_{cP}}\right) = \sum_{P} \xi_{cP} H_{ex,c}^{P} \quad .
\end{aligned} \tag{5.6}$$

By inverting Eq. 5.6 it is possible to separate the fine diversity into an inter-sectorial and an intra-sectorial contribution; moreover, by implementing the boundedness of the Shannon's entropy function, the same equation also proves the validity of the condition $H_{ex,c} \geq H_{ex,c}^{CG}$ (the equality holding if all the macro-categories contain only a single non-null export). The relation between the different aggregation levels is analyzed in Figure 21a, which clearly highlights the tendency by the average intra-sectorial contribution to grow along with the fine diversity. Changing perspective, Figure 21b relates the same contribution to the overall degree of development of a country, represented by the monetary value of its export basket. As shown, the countries with the most relevant baskets (in terms of both the monetary value and the

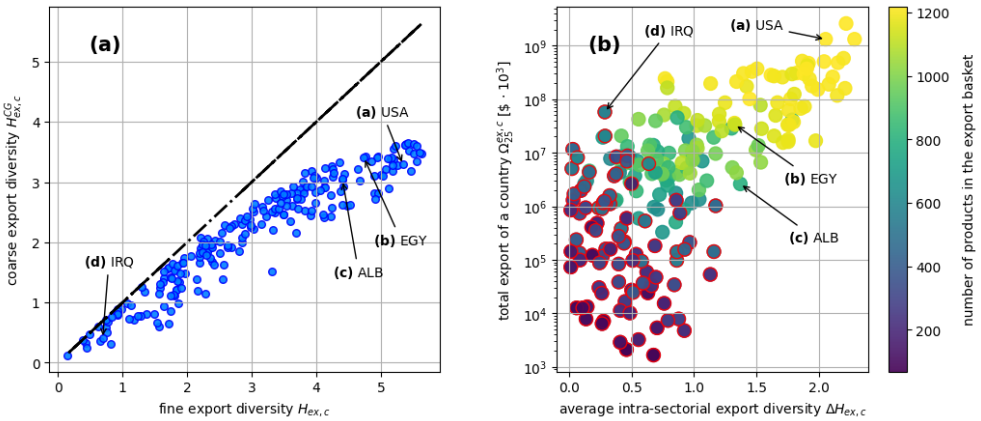

Figure 21: relation between the diversities associated to different levels of aggregation. Panel **(a)** describes the relation between the fine diversity $H_{ex,c}$ and the coarse one $H^{CG}_{ex,c}$, while panel **(b)** provides a comparison between the total export and the average intra-sectorial diversity $\Delta H_{ex,c}$. In panel **(b)**, the markers circled in red describe the region occupied by the poverty trap area. In both plots, some arrows indicate the position of United States (USA), Egypt (EGY), Albania (ALB) and Iraq (IRQ).

number of exported products) present the highest internal organization degrees within their productive sectors, a feature which clearly distinguishes them from the countries which lie inside the poverty trap area. Furthermore, the clear distinction between the four representative countries United States, Egypt, Albania and Iraq allows to consider the intra-sectorial contribution $\Delta H_{ex,c}$, as well as $H_{ex,c}$, as a reliable measure of the economic efficiency.

5.4 Dynamical characterization of the network of countries

The chapter concludes with the presentation of a novel macro-economic characterization of the global network of countries, which embodies the information provided by the application of both the diversity measure and the dynamic model. The results presented in Section 5.3 have highlighted the usefulness of the diversity as an estimator of the economic efficiency, providing a neat distinction between the group of most developed countries and the group of the countries which are confined inside the poverty trap area. Therefore, by adding a supplemental dimension, entirely based on the export activity, to the economic narrative, the diversity naturally lends itself to the cooperation with other, traditional macro-economic indicators in the context of the assessment of the overall degree of development of a country. Following the example of [14], Figure 22 defines an economic benchmark by comparing the diversity and the gross domestic product per capita (GDPpc) related to all the countries in the network, in year 2020. The vertical dimension obviously distinguishes the countries according to the wealth of the respective society; on the other hand, in view of the

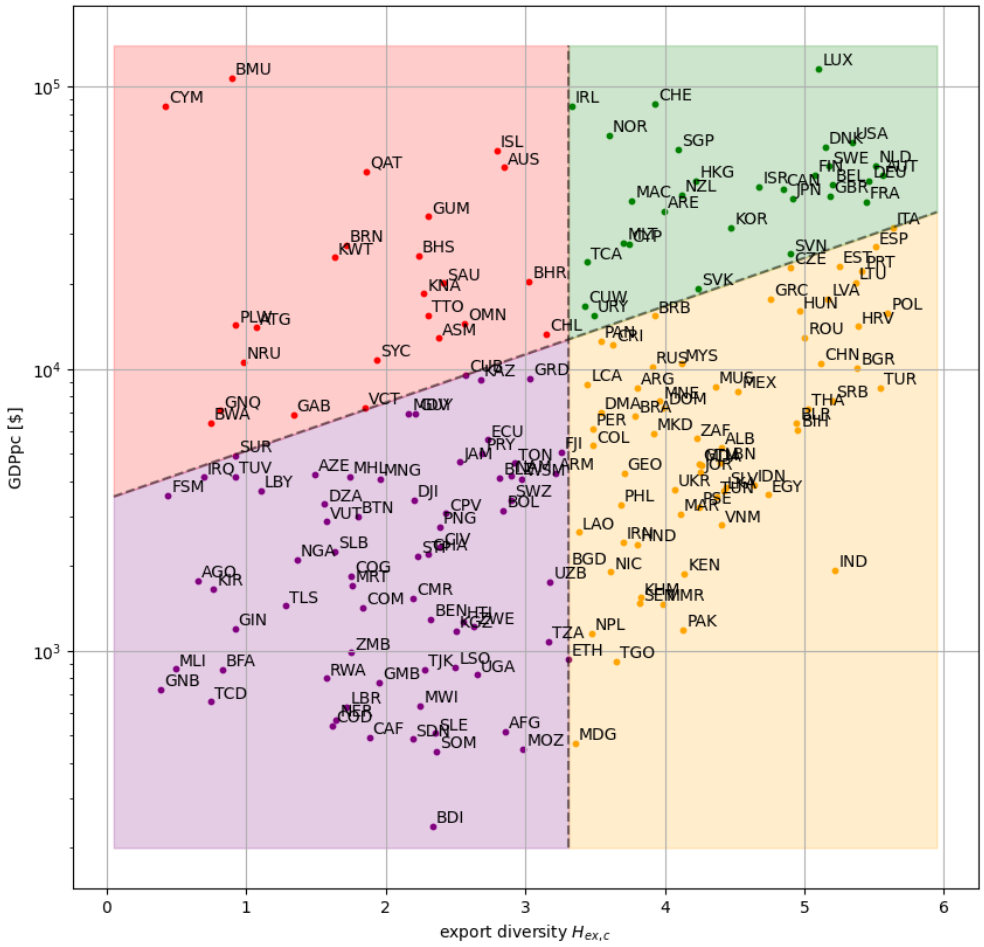

Figure 22: macro-economic benchmark relating the gross domestic product per capita (GDPpc) with the diversity of the countries. The four regions highlighted in the plot are defined by means of a vertical dotted line, indicating the median of the diversity distribution, and an oblique one, representing the exponential fit of all the points.

discussion presented in Section 5.3, the horizontal dimension evaluates not only the level of stability and the exposure to the fluctuations of the global market but also the economic efficiency of the countries considered. The oblique dotted line results from the exponential regression of all the points and highlights the positive correlation between the GDPpc and the diversity. The vertical dotted line, instead, stands in correspondence of the median of the diversity distribution and defines, along with the regression line, the four colored regions highlighted in the plot.

Figure 23, on the other hand, presents the evolution across the period 1995-2020 of the position on the same benchmark of all the countries for which a complete set of

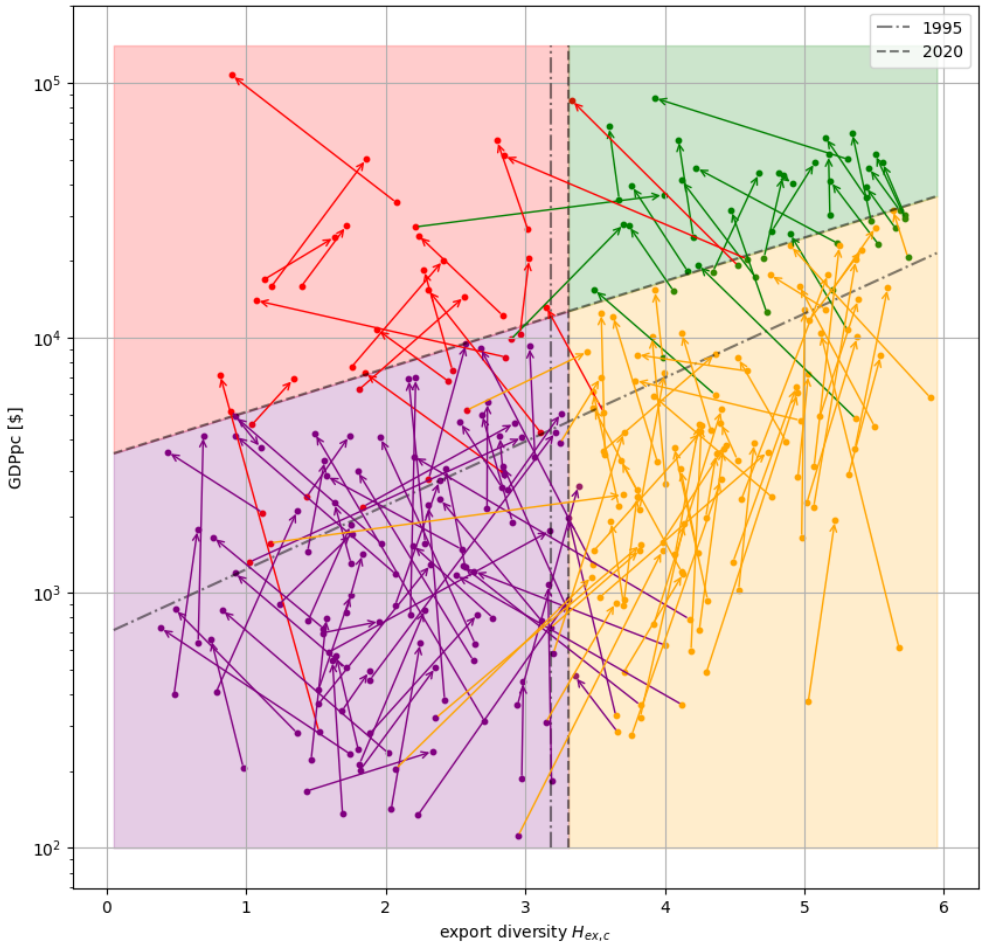

Figure 23: evolution of the countries on the macro-economic benchmark across the period 1995-2020. The limiting points of the evolution of each country refer to the region occupied during the respective year. The four regions highlighted in the plot are related to year 2020.

records exists (for example, as explained in Chapter 2, some early records are missing in the case of a few "crippled" countries). The color related to each point refers to the economic condition (in terms of the regions of the benchmark) of the respective country in a certain year. First of all, mainly due to the inflationary growth of the prices, all the countries have been interested by an upward movement towards high-GDPpc regions. Moreover, the countries which have at some point passed through the right part of the plot have experimented either an exclusively vertical evolution or a regression towards less developed regions, demonstrating the substantial stability of the top diversity standard across the whole period; on the other hand, the evolution

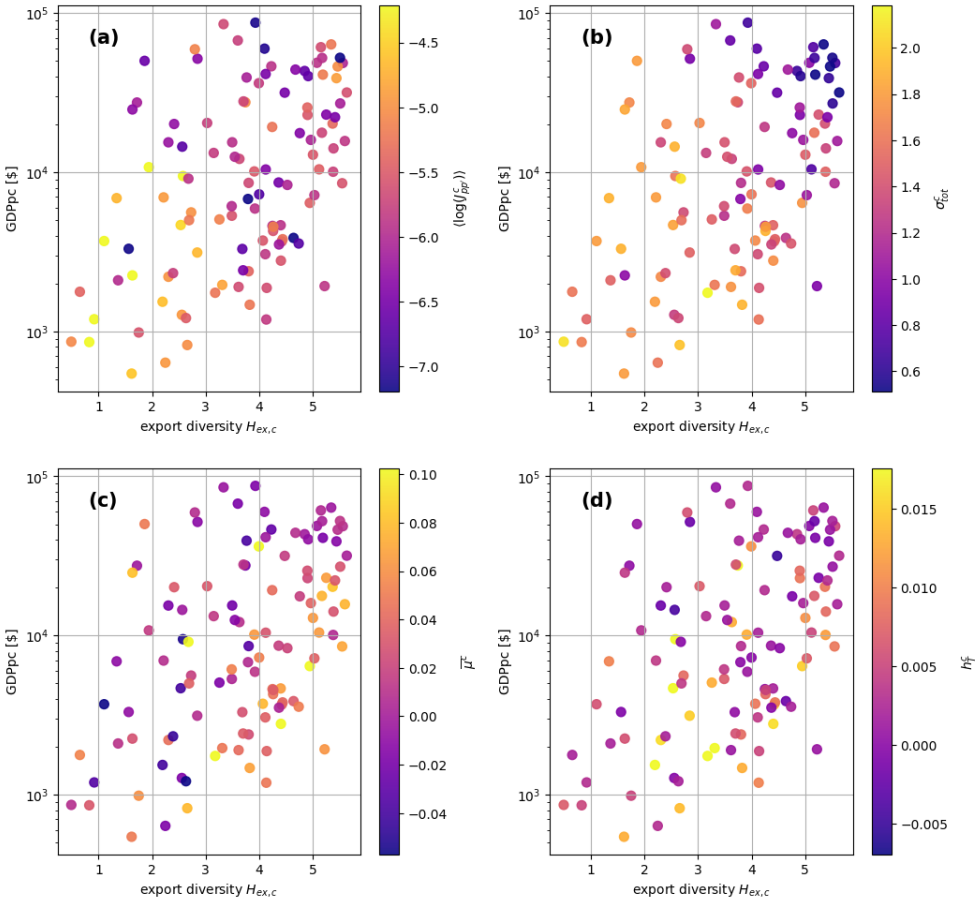

Figure 24: dynamical characterization of the macro-economic benchmark through the implementation of the calibrated parameters $\langle J^c_{pp'} \rangle$ **(a)**, σ^c_{tot} **(b)**, $\overline{\mu}^c$ **(c)** and h^c_T **(d)**.

of the countries on the left side of the plot has clearly been much more chaotic and unpredictable. Given the limited number of countries which have managed to evolve towards a more developed region, it is easy to observe the general stagnation which has interested the international scenario during the last 25 years.

Then, the plots in Figure 24 enrich the characterization of the macro-economic benchmark by finally introducing the parameters resulting from the calibration procedures. As shown in Figure 24a and Figure 24b, the average rate of transfer of resources and the total noise term, respectively defined as

$$\langle J^c_{pp'} \rangle \equiv \frac{\sum_{p,p'} G^c x^c_p |C^c_{pp'}|}{(N^c)^2} \quad \text{and} \quad \sigma^c_{tot} \equiv \sqrt{(\sigma^c)^2 \tau^c + (\sigma^c_0)^2} \quad ,$$

provide the clearest distinctions between developed and underdeveloped countries. In

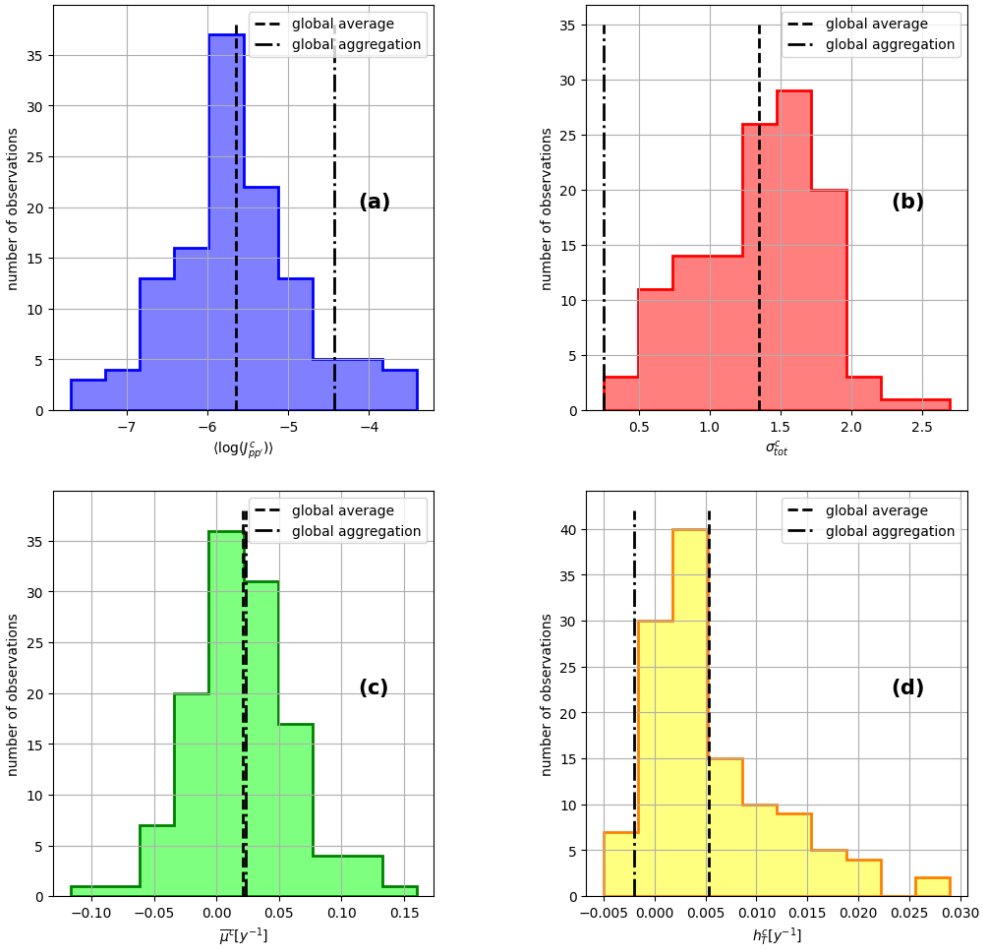

Figure 25: histograms of the global distributions of the calibrated parameters $\langle J^c_{pp'} \rangle$ **(a)**, σ^c_{tot} **(b)**, $\overline{\mu}^c$ **(c)** and h^c_T **(d)**. The dashed lines indicate the average global values of the parameters and the ones related to the aggregation of the exports at the global level.

the first case, a diagonal gradient from the bottom left corner to the top right one is clearly perceivable and highlights the general weakness of the transfer mechanisms related to the most developed countries. This result is conceptually in agreement with the above-average stability degree which has characterized these countries across the period of interest, due to the absence of relevant upheavals of the respective productive sectors. Conversely, a strong redistribution of resources is obviously peculiar of rapidly evolving economies, which are typical of underdeveloped or developing countries. While the distribution of the values in the top right corner is quite homogeneous, the bottom left one presents a considerable number of outliers. An equivalent gradi-

ent is observable in an even clearer manner inside the total noise term plot. In this case, the combination of strong inner stability and resistance against the turbulence of the global market neatly detaches the most developed countries from the rest of the network; conversely, the countries in the other regions of the plot assume consistently higher values of the parameter.

Although the economic depiction provided by Figure 24c and Figure 24d is much less clear, it is still possible to recognize in both cases a high homogeneity degree in the most developed corner. The drift parameter $\overline{\mu}^c$ describes the average, independent growth rate of an export basket and can be in principle controlled, as well as the transfer parameter G^c, through the application of particular investment policies: therefore, the consistently lower (often even negative) values assumed by the parameter in the top right corner indicate that the injection of fresh resources in the productive infrastructure starts to be severely hindered once a country reaches a relatively optimal degree of development. In agreement with this interpretation, the rapidly developing economies in the bottom right corner are characterized by consistently higher values of the parameter.

The last plot shows the contributions to the growth of the baskets (Eq. 3.31) originating from the transfer mechanism of the dynamic model. In agreement with the results presented in Section 5.2, the high degree of stability typical of the most developed countries tends to nullify any potential positive contribution generated by such mechanism. The remaining regions of the plot present much more variegated distributions of the parameter and no clear pattern can be recognized; however, the occasional occurrence of exceptionally high values of h_T^c in the least developed corner, in conjunction with the patterns of Figure 24a and Figure 24b, indicates that a relevant positive transfer contribution can originate from the combination of a high average transfer rate and high fluctuations levels.

Lastly, the histograms in Figure 25 provide an additional insight concerning the global distribution of the calibrated parameters. By means of vertical dotted lines, the plots also indicate the average global values and the ones related to the aggregation of the exports at the global level: in agreement with the conceptual meaning of the parameters, the World correctly behaves as a very developed country if the transfer contribution and the total noise term are considered and presents intermediate properties in the case of the average transfer rate and the drift term.

6 Causal relation between diversification and wealth

Overturning the perspective which has subtended the discussion up to this point, the final chapter of this work aims to expand the analysis of the layer of countries by exploring the predictive potential of both the dynamic model and the complexity measures. While in chapter 5 the application of these methods has been limited either to the reproduction, realistic (Section 5.1) or counterfactual (Section 5.2), of the export baskets or to the static characterization of the economic development of the countries in a given year (Section 5.4), the discussion in the sections that follow will investigate the causal connection between the evolution of the diversity of the baskets and the macro-economic narrative, in order to lay the foundation for the construction of meaningful predictions about the wealth prospects of the countries. This task, extensively pursued in the economic complexity field (see, for example, [40], [41] and [42]), turns out to be intriguing for at least two reasons. First of all, it allows to expand the scope of the Shannon's entropy function by exploring its potential not only as an additional, uncorrelated dimension in the characterization of the layer of countries (see Figure 22) but also as a useful forecasting tool in tight correlation with other macro-economics indicators. Second of all, as will be proved in Section 6.2, the same task is optimally fulfilled if also the information concerning the import activity of the countries is embodied in the analysis and offers, then, the opportunity to investigate the structure of the related bipartite network and its relation with the export one.

This last topic is treated in the first section of the chapter. Afterwards, the subsequent section presents the application of the *convergent cross-mapping method* (CCM), which allows the assessment of the dynamical correlation between two quantities; this section is further subdivided into two paragraphs, dedicated to the analysis of, respectively, the empirical scenario and a synthetic one resulting from the implementation of the dynamic model.

6.1 Analysis of the import bipartite network

The procedure employed in the construction of the import bipartite network is completely analogous to the one presented in Chapter 2 for the export one. First of all, the records on the exporter's axis of the 4-digits coarse-grained version of the database have been aggregated in separated groups related to just as many importing countries; subsequently, the singularities which occasionally affect the resulting database have been treated depending on the purpose of the analysis. The final collective records are denoted with $Y^c_{p,n}$, adopting the same conventions seen in the previous chapters.

As in the export case, the structure of the import network can be effectively represented by the respective bi-adjacency matrix, depicted in Figure 27 along with the related survival function's diagram. A quick comparison with Figure 1 in Chapter 2 reveals a notable distinction in the distributions of the links which constitute the two networks: in the import case, the average node of both the layers is the source of a much higher number of links and the survival function presents a consequent pronounced belly shape. On the other hand, a remarkable analogy between Figure 1b and Figure 27b is the exclusive occupation of the upper region of the plots by the group of most developed countries. This last property may seem counter-intuitive, especially in the perspective of the classic economic theory. Adopting the latter, the development process of the economy of a country should be characterized by an increasing self-sufficiency degree in the production of all the necessary resources, which would obviously reduce the need of weaving import trade agreements with other countries. From this point of view, a very developed country should present a scarce import basket, at least if compared to the basket of an underdeveloped country with a comparable population. However, this statement is clearly refuted not only by the import survival function' diagram but also by the empirical evolution of the import baskets from United States, China and Russia (Figure 26), which present distributions of the imports which are pretty similar, in monetary terms, to the ones observed in their ex-

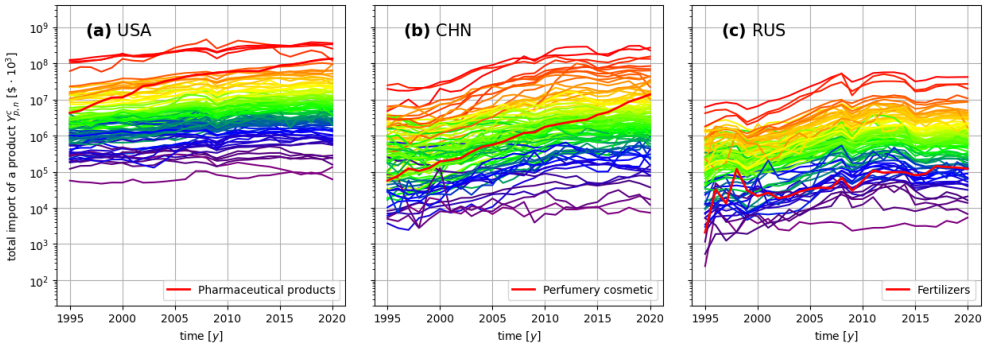

Figure 26: empirical evolution across the period 1995-2020 of the import baskets of United States (USA, **(a)**), China (CHN, **(b)**) and Russia (RUS, **(c)**) at the 2-digit aggregation level. The wavelength of the color assigned to each product reflects its rank (Eq. 2.1) inside the respective basket. For each country, the product interested by the most relevant growth (in relative terms) is marked with a thick red line.

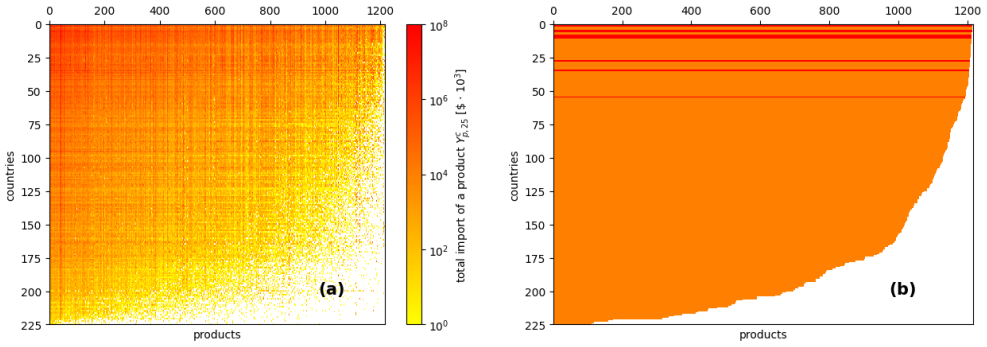

Figure 27: bi-adjacency matrix **(a)** and survival function's diagram **(b)** related to the import bipartite network in year 2020. In panel **(b)**, a set of darker lines highlights the position of the top-10 section of the GDP ranking.

port counterparts in Figure 2. The origin of this discrepancy can be easily attributed to the strong globalization process which has interested the global economy in the last 60 years. The most relevant outcome of this process, at least on the economic plane, is indeed the creation of a global productive system, nowadays spread across a multitude of different countries. Considering this situation, it seems natural to deem the degree of integration inside such system (which can be estimated through the monetary assessment of the baskets and the analysis of the links of the bipartite networks) as a reliable measure of the competitiveness on the global market and, ultimately, of the degree of development of a country.

In each of the plots in Figure 26, the products interested by the most relevant import growth (in relative terms) across the period 1995-2020 are marked with a thick red line. In all the other cases, the assignment of the wavelength has followed the same criterion adopted for the export baskets (see the related discussion in Chapter 2). The sharpness of the resulting rainbow effects helps to assess the degree of stability which has characterized the aggregated demand of a given country during the whole period. Interestingly, for each of the three countries considered, the stability of the aggregated demand seems to match the stability of the productive sector. A ranking of the products contained in a given import basket can be defined, just like in the export case, in order to describe the hierarchy of the distribution: matching the export definition, the rank of product p is given by the expression

$$y_p^c \equiv \frac{1}{5} \sum_{n=0}^{T} \frac{Y_{p,n}^c}{\Omega_n^c} \quad , \tag{6.1}$$

with

$$\Omega_n^{im,c} \equiv \sum_{p=0}^{N_n^c} Y_{p,n}^c \quad .$$

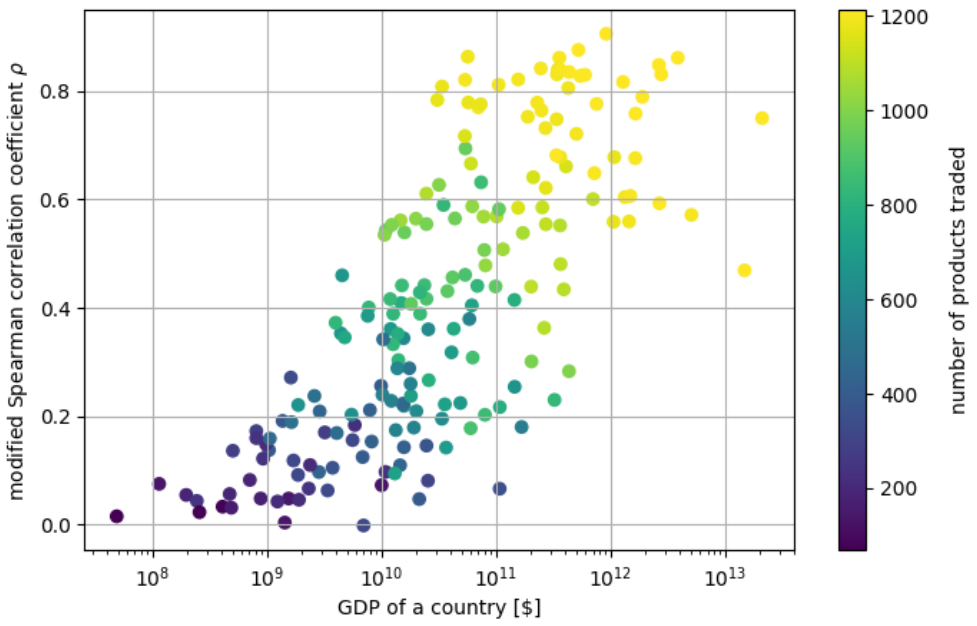

Figure 28: relation between the GDP of the countries and the overlap degree between their export and import baskets, estimated through the modified version of the Spearman's coefficient ρ_S.

Employing once again the versatility of the Spearman's correlation coefficient, the export and the import rankings can be confronted in order to determine the origin of the aggregated demand of the countries. While, as explained before, the import activity of a very developed country can be connected with the needs of the global productive system, it seems reasonable to assume that, conversely, the import trade agreements signed by an underdeveloped country generally reflect the internal needs of its society. This statement is confirmed by Figure 28, which relates the coefficients between the rankings $\vec{x}^c$ and $\vec{y}^c$ with the GDP of each country in year 2020. In order to account for the correlation boost which arises from the relative scarceness of many baskets, the Spearman's coefficients have been multiplied by the corrective factor N^c/N^{WRD}, where the numerator indicates the number of products traded in both directions by country c during the last five-years period. As shown, less developed countries are related to consistently lower coefficients, which highlight the expected poor overlap between their productive capabilities and the imports which define their demand.

At this point, it is interesting to verify the possible agreement of the outcomes of the iterative algorithm described in Section 4.1 with the scenario just depicted. The procedure of construction of the algorithm is obviously equivalent to the one related to the export scenario and starts with the definition of the bare complexity measures. In the import case, the bare diversity and ubiquity are respectively defined as

$$H_{im,c}^{(0)} = -\sum_{p=0}^{N^c} \chi_{cp}^{(0)} \log\left(\chi_{cp}^{(0)}\right) \quad \text{and} \quad H_{im,p}^{(0)} = -\sum_{c=0}^{N^p} \psi_{cp}^{(0)} \log\left(\psi_{cp}^{(0)}\right) \quad , \tag{6.2}$$

where the quantities

$$\chi_{cp}^{(0)} \equiv \frac{Y_p^c}{\sum_{p'}^{N^c} Y_{p'}^c} \quad \text{and} \quad \psi_{cp}^{(0)} \equiv \frac{Y_p^c}{\sum_{p'}^{N^p} Y_{p'}^c} \tag{6.3}$$

denote the bare shares of the products.

Starting from these, the line of reasoning which leads to the formal definition of the iterative algorithm is again related to the need of re-weighing the bare shares, according to the intertwined nature of the bipartite network described in Section 4.1: substantially, a widely imported product should carry a reduced weight in the evaluation of the diversity of an import basket and vice versa. The algorithm, then, takes a form which is completely analogous to Eq. 4.4:

$$\begin{cases} H_{im,c}^{(k+1)} = -\sum_{p=0}^{N^c} \chi_{cp}^{(k)} \log\left(\chi_{cp}^{(k)}\right) \\ H_{im,p}^{(k+1)} = -\sum_{c=0}^{N^p} \psi_{cp}^{(k)} \log\left(\psi_{cp}^{(k)}\right) \end{cases} , \tag{6.4}$$

with the shares at the k-th step defined as

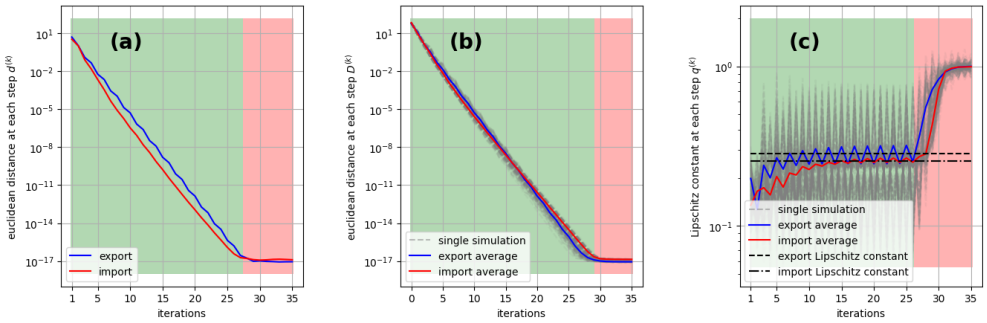

Figure 29: convergence properties of the iterative algorithm in both the export and the import scenario: Euclidean distance between consecutive steps $d^{(k)}$ (Eq. 4.8) **(a)**, Euclidean distance from the empirical fixed point $D^{(k)}$ (Eq. 4.9) **(b)** and Lipschitz constant $q^{(k)}$ related to the map ϕ (Eq. 4.7) **(c)**. In panels **(b)** and **(c)**, the colored lines highlight the evolution of the mean Euclidean distances extracted from two distinct sets of 100 simulations.

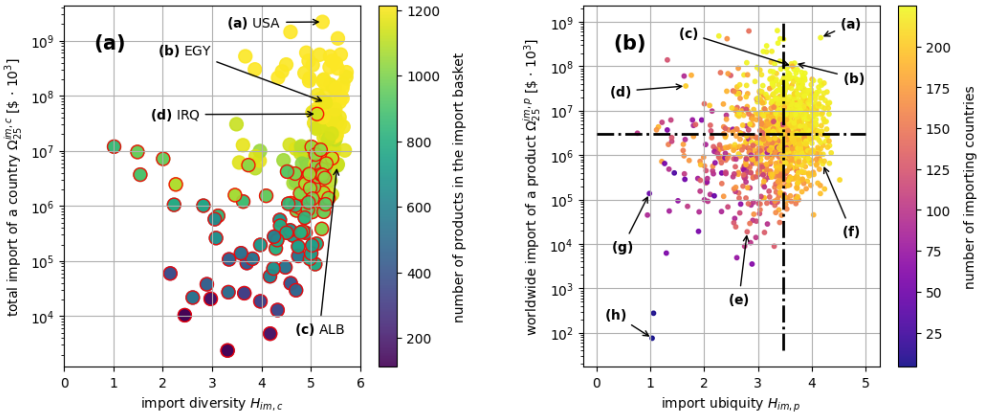

Figure 30: relation between the final complexity measures and the total import of each node of the import bipartite network, in year 2020. In panel **(a)**, the markers circled in red describe the region occupied by the poverty trap area, while the arrows indicate the position of the four representative countries United States (USA), Egypt (EGY), Albania (ALB) and Iraq (IRQ). In panel **(b)**, the black dashed lines refer to the medians of the quantities represented on the axes and subdivide the global market in four regions, while the arrows indicate the position of non crude oils **(a)**, insulated electric conductors **(b)**, aircrafts and spacecrafts **(c)**, photographic laboratory equipment **(d)**, hand-woven tapestries **(e)**, paints and varnishes **(f)**, cobalt ores **(g)** and phosphides **(h)**.

$$\chi_{cp}^{(k)} = \frac{Y_p^c \; f\left(H_{im,p}^{(k)}\right)}{\sum_{p'=0}^{N^c} Y_{p'}^c f\left(H_{im,p}^{(k)}\right)} \quad , \qquad \psi_{cp}^{(k)} = \frac{Y_p^c \; g\left(H_{im,c}^{(k)}\right)}{\sum_{c'=0}^{N^p} Y_p^{c'} g\left(H_{im,c}^{(k)}\right)} \tag{6.5}$$

and the re-weighing functions as in Eq. 4.6. Regardless of the number of steps implemented, the algorithm 6.4 can be mathematically represented by the continuous map ϕ (Eq. 4.7), acting from a compact and convex real subset S onto itself; the properties of the shares in Eq. 6.5 and S allow to apply the Brouwer's theorem and to state the existence of at least one fixed point in the inner region of S itself.

Figure 29 demonstrates the uniqueness and the stability of the empirical fixed point by showing the evolution across multiple iterations of the Euclidean distance between consecutive steps **(a)** (Eq. 4.8), the Euclidean distance from the empirical fixed point itself **(b)** (Eq. 4.9) and the Lipschitz constant **(c)** (Eq. 4.10), each of which is confronted with its respective export counterpart. The curve in Figure 29a has been obtained by setting the empirical bare complexity measures (Eq. 6.2) as initial conditions of the algorithm; on the other hand, the average curves in Figure 29b and Figure 29c have been extracted from distinct sets of 100 simulations, each related to a different point randomly selected in the inner region of S. As shown,

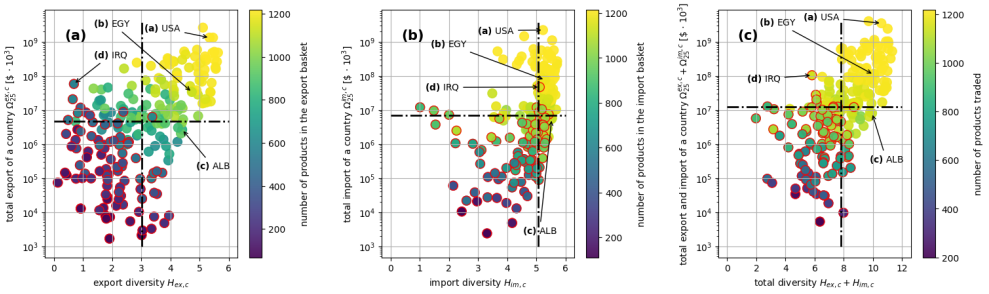

Figure 31: relation between the export diversity $H_{ex,c}$ and the total export $\Omega_{25}^{ex,c}$ **(a)**, between the import diversity $H_{im,c}$ and the total import $\Omega_{25}^{im,c}$ **(b)** and between the total diversity $H_{tot,c} \equiv H_{ex,c} + H_{im,c}$ and the total sum of export and import **(c)**, in year 2020. In each panel, the black dashed line indicate the medians of the distributions of the quantities represented on the axes, while the markers circled in red describe the region occupied by the poverty trap area. A set of arrows indicates, in every panel, the position of the four representative countries United States (USA), Egypt (EGY), Albania (ALB) and Iraq (IRQ).

the export and the import curves converge towards the double precision limit at very similar rates; the only perceivable differences between the two cases consist in a slightly faster approach of the import measures towards the respective empirical fixed point and in the characterization of the import algorithm as a slightly stronger contraction, as witnessed by the lower limiting Lipschitz constant ($q_{im} \approx 0.26$, while $q_{ex} \approx 0.28$).

Finally, the plots in Figure 30 offer a representation of the empirical fixed point reached by the import algorithm in year 2020, separating the layer of countries from the one of the products. In both plots, the color assigned to each marker reflects the number of links which branch out from the related node. Confronting Figure 19a and Figure 30a, a clear distinction between the dispositions of the markers immediately stands out: differently from the export bipartite network, the import one is characterized by a substantial homogeneity in the distribution of the respective diversity values, which does not allow a clear separation between the macro-categories of national economies defined in Figure 7. In particular, this fact is highlighted by the positions occupied by the representative countries United States, Egypt, Albania and Iraq, which, despite being well distinguished on the total import dimension, share extremely similar diversity values. More importantly, the absence in Figure 30a of a clear separation between the corner of most developed countries and the poverty trap area, whose markers share once again a red edge, negates the possibility to employ the import diversity, in place of the export one, as an effective measure of the economic efficiency of a country. Another difference, not immediately perceivable from the plot, consists in the relatively poor overlap between the export diversity ranking and the import one, as witnessed by the intermediate value assumed by the related Spearman's coefficient ($\rho_S = 0.52$). For example, this feature can be appreciated by comparing the top-10 and the bottom-10 sections of the two rankings, represented in Table 2.

Conversely, the analysis of Figure 30b does not reveal significant differences with

Top-10 Diversity Ranking				
	Export Layer		**Import Layer**	
Rank	**Country**	**Diversity**	**Country**	**Diversity**
1	Italy	5.63	*Poland*	5.72
2	*Poland*	5.60	Russia	5.71
3	*Austria*	5.56	Bosnia Herzegovina	5.68
4	Turkey	5.55	Serbia	5.66
5	Spain	5.51	*Austria*	5.65
6	Netherlands	5.51	Romania	5.64
7	Germany	5.46	Indonesia	5.63
8	*France*	5.44	Denmark	5.60
9	Portugal	5.41	*France*	5.60
10	Croatia	5.39	Lithuania	5.59
Bottom-10 Diversity Ranking				
	Export Layer		**Import Layer**	
Rank	**Country**	**Diversity**	**Country**	**Diversity**
216	Chad	0.75	British Virgin Islands	2.82
217	Botswana	0.74	*Niue*	2.61
218	Iraq	0.69	Tokelau	2.44
219	Angola	0.65	Bermuda	2.26
220	South Sudan	0.62	Saint Lucia	2.23
221	Mali	0.49	Tuvalu	2.15
222	Federated State of Micronesia	0.44	Gibraltar	2.01
223	*Cayman Islands*	0.42	*Cayman Islands*	1.54
224	Guinea-Bissau	0.39	Liberia	1.49
225	*Niue*	0.14	Marshall Islands	1.02

Table 2: Top-10 and bottom-10 sections of the export and the import diversity rankings. The countries appearing in analogous sections of the two rankings are indicated in italic.

Figure 19b in terms of the global distribution of the markers. Apart from this aspect, in the import case the median of the ubiquity distribution is considerably shifted towards the right section of the plot and a horizontal gradient related to the number of importing countries is clearly perceivable. A set of arrows indicates the position of the same products highlighted in Figure 19b: with the only exception of hand-woven tapestries **(e)**, the selected products are found in market regions which are equivalent to their export counterparts. Despite the high degree of overlap between the regions, the relatively low Spearman's coefficient ($\rho_S = 0.48$) between the export and the import rankings (both briefly represented in Table 3) demonstrates the overall lack of correlation between the distributions of the two ubiquity measures.

While the import diversity alone has proved useless as an economic efficiency indicator, a comparison between the dispositions of the markers in Figure 19a and Figure 30a (the former slightly concave, the latter slightly convex) suggests that the cooperation between the two diversities can provide an improvement in this sense. Figure 31 realizes their unification by relating, in panel **(c)**, the sum between the two measures with the overall degree of development of the countries. For the sake of symmetry with

Top-10 Ubiquity Ranking				
	Export Layer		**Import Layer**	
Rank	**Product**	**Ubiquity**	**Product**	**Ubiquity**
1	Copper waste and scrap	3.89	Iron or steel barbed wire	4.51
2	Hydraulic cements	3.87	Rubber inner tubes	4.36
3	Plastic waste, parings and scrap	3.74	Organic solvents and thinners	4.33
4	Frozen fishes	3.73	Machinery for mineral working	4.32
5	Sugar confectionery	3.72	Hypochlorites and hypobromites	4.32
6	Residues of cereals working	3.72	Machinery for seed working	4.32
7	Bakers wares	3.71	Soap preparations	4.32
8	Perfumery and pharmacy plants	3.71	Iron angles and shapes	4.30
9	Jams and fruit jellies	3.70	Iron bars and rods	4.30
10	Aluminium waste and scrap	3.69	Ceramic building bricks	4.29
Bottom-10 Ubiquity Ranking				
	Export Layer		**Import Layer**	
Rank	**Product**	**Ubiquity**	**Product**	**Ubiquity**
1232	Glazed ceramic flags and paving	1.03	Tin ores and concentrates	1.27
1233	*Coal and water gas*	1.01	Aluminium ores and concentrates	1.22
1234	Festive articles	1.00	Ground-nut oils	1.21
1235	Residues of ground-nut oil	1.00	Cast and rolled glass	1.15
1236	*Articles of precious metal wares*	0.92	*Articles of precious metal wares*	1.12
1237	Pigs bristles and hair	0.91	Word-processing machines	1.06
1238	Artificial flowers and fruit	0.64	Phosphides	1.03
1239	Cobalt oxides and hydroxides	0.42	*Cobalt ores and concentrates*	0.98
1240	Jute yarn	0.29	*Coal and water gas*	0.95
1241	*Cobalt ores and concentrates*	0.25	Lignite	0.76

Table 3: Top-10 and bottom-10 sections of both the export and the import ubiquity rankings. The products appearing in analogous sections of the two rankings are indicated in italic.

the other plots, the latter is estimated through the sum of the monetary amounts of export and import of the countries; however, this quantity could be substituted by both the single amounts on their own without causing relevant alterations to the structure of the plot, due to the extremely high correlation between their distributions. Again in panel **(c)**, the number of products contained in the union of the export and the import baskets is the criterion for the assignment of a marker's color; in this perspective, the set of countries can be represented from a topological point of view as a layer of a hybrid bipartite network, whose links reflect the presence of a certain product inside either the export or the import baskets.

Remarkably, the disposition of the markers in the third plot clearly shows that the unification of the export and the import diversities, in addition to supplying a more complete description of the commercial activity of the countries, is able to generate a novel measure which appears to be more strongly correlated with the overall degree of development of the countries. More importantly, the same measure also proves to be a better economic efficiency indicator than the single export diversity, as it induces a clearer detachment of the poverty trap area from the region of developed countries. In particular, the median of the total diversity distribution works as a simple and efficient delimiter between the two regions, as it manages to minimize the degree of contamination between them.

As a last remark, it is important to state the excellent overlap degree ($\rho_S = 0.94$) between the rankings related to the diversities represented in panels **(a)** and **(c)**: because of this property, despite the improvement in the assessment of the economic efficiency, the total diversity does not provide any additional contribution to the economic characterization of the countries and no clear advantage can be recognized in its implementation, in place of the single export diversity, into the analysis presented in Section 5.4.

6.2 Convergent cross-mapping method

The theoretical paradigm which assesses the dynamical correlation between two quantities, denominated *convergent cross-mapping method* (CCM) by its main developer Sugihara, has been introduced in [30] in the context of the dynamical analysis of complex ecosystems. After a quick description of the purpose of the method and of the conceptual properties which distinguish it from other similar paradigms, the authors procede to successfully apply it to a series of frequently analyzed case studies, which ranges from systems driven by unidirectional couplings exercised by external forcing entities (like the interaction between the temperature of the sea and the population of fishes) to systems governed by bidirectional couplings, both of competitive nature (like systems of predators and preys) and cooperative nature (like systems of species in a symbiosis relation). However, the full generality of the framework which supports the development of the method opens the possibility of its employment in branches of research which are completely detached from the biological one. As will be clarified later, the main element of distinction between CCM and other paradigms found in the literature is its ability to supply necessary conditions for causality, which allow a clearer characterization of the correlation evidence.

The first paragraph of this section is devoted to a brief informal representation of the theoretical background of CCM and of its main purposes. Then, the subsequent paragraphs will present the results of the application of CCM both in the empirical macro-economic scenario and in a simulated one, generated by means of the dynamics model. The conceptual origin of CCM and the technical details concerning its theoretical background will be extensively discussed in Appendix A.

6.2.1 Illustration of the method

The theoretical result which supports the development of CCM is provided by a modified version of Takens's delay embedding theorem, originally formulated in [43]. From its perspective, the evolution in time of the system of interest is assumed to take place on an attractor manifold M characterized by a fractal structure. Then, the theorem sets the conditions for the construction of an additional peculiar manifold, embedded in a real E-dimensional space and diffeomorphic to M. An upper and a lower limit to the range of feasible values of E are respectively induced by the actual dimension of M and by its fractal one. Specifically, the embedded manifold is defined from the set of time-lagged E-dimensional vectors built from the paths of a smooth function on M. For example, if function X satisfies the condition of the theorem, the point reached by the system at the generic instant t on a given trajectory on M is identified with a vector of the form $(X(t), X(t+\tau), ..., X(t+(E-1)\tau))$, where τ is an arbitrarily selected (positive) time lag. Obviously, different trajectories on M are related to different paths of X and to different time-lagged vectors on the embedded manifold, hereafter indicated with M_X. If the additional function Y satisfies in turn the condition of the theorem, the analogous diffeomorphic manifold M_Y can be defined from the respective time-lagged vectors. In this case, through a simple composition of the diffeomorphisms, the time-lagged vectors on M_X can be exactly matched to their equivalent ones on M_Y and vice versa (the two functions are said to "cross-map" each other):

$$(X(t), X(t+\tau), ..., X(t+(E-1)\tau)) \leftrightarrow (Y(t), Y(t+\tau), ..., Y(t+(E-1)\tau)) \quad , \quad (6.6)$$

$\forall t \geq 0$ and for all trajectories. As the intuition suggests, this result reflects the presence of an effective causal relation between the two functions, which evolve on M following bidirectionally coupled paths. Conversely, if the condition of the theorem is satisfied only by, say, X, the matching between the sets of time-lagged vectors can be accomplished only in one direction,

$$(X(t), X(t+\tau), ..., X(t+(E-1)\tau)) \rightarrow (Y(t), Y(t+\tau), ..., Y(t+(E-1)\tau)) \quad , \quad (6.7)$$

because the bijective correspondence between M and M_Y cannot be stated with certainty: in this second case, X does not exercise an effective causal influence on the dynamics of Y and, thus, the coupling between their paths on M is unidirectional.

Taking for granted the fractal nature of M (for example, such a feature is always adequate for the description of chaotic systems), CCM allows to verify a couple of necessary conditions for the validity of the above identifications in the case of two selected observables of the system and, consequently, for the presence of a causal relation between them. This purpose is practically pursued through the implementation of an algorithm for the analysis of the empirical time series, which, according to the following chains of steps, matches the sets of time-lagged vectors and provides a reconstruction of the single records:

$$(X(t_n), ..., X(t_n+(E-1)\tau)) \rightarrow (Y(t_n), ..., Y(t_n+(E-1)\tau)) \rightarrow Y(t_n)$$

and

$$(Y(t_n), ..., Y(t_n+(E-1)\tau)) \rightarrow (X(t_n), ..., X(t_n+(E-1)\tau)) \rightarrow X(t_n) \quad , \quad \forall n \quad ; \quad (6.8)$$

at this point, the dynamical correlation between the observables involved can be estimated through the matching degrees between the empirical time series and their reconstructed counterparts. The first necessary condition that can be verified by means of CCM is the excellent matching degree that should be detected between the time series related to, say, observable X if a causal influence from X itself on the dynamics of Y is effectively present. Then, considering again this scenario, the second necessary condition that can be inspected is the monotonic convergence of the same matching degree towards a maximum asymptotic value as longer portions of the empirical time series get involved into the analysis. Indeed, if X exercises an effective influence on the dynamics of Y, the filling of the trajectory on M_Y resulting from the growth of the empirical time-lagged vectors' population should progressively improve the estimation of the records of X itself, leading to a monotonic increase of the matching degree; moreover, this result should be verified regardless of the criterion adopted for the addition of further empirical records into the analysis. These necessary conditions will be both investigated in all the case studies presented in the subsequent paragraphs.

As a last remark, it is important to state the relevance of all the limiting factors (including observational errors, process noise and limited length of the time series)

which typically affect the analysis of the experimental data: for this reason, all the conclusions that can be drawn from the final results provided by CCM should be always interpreted in a probabilistic perspective.

6.2.2 Empirical dynamical correlation

This paragraph presents the evaluation through CCM of the dynamical correlation between the diversity measures collected in Figure 31 and the GDP, with the purpose of determining which measure exhibits the tightest connection with the overall degree of development of a country. Following the choice adopted in [30], the matching degree between the empirical and the reconstructed time series has been estimated in every case through the Pearson's correlation coefficient. Moreover, the manifolds resulting from Takens's theorem have always been embedded into a minimal two-dimensional space, with time-lagged vectors of the form $(X(t_n), X(t_n + 1))$ (selecting $\tau = 1$ yr): in other words, the value of E has been set to 2 in every case study considered.

The choice of the E-value is strongly influenced by the length of the time series (denoted with L in all the subsequent plots) which are available for the analysis. A higher E-value allows to connect in a tighter manner the dynamics of the observables involved and leads to more reliable results from the interpretative point of view. For this reason, in an optimal scenario, the analysis should consider the implementation of increasingly larger embedding spaces up until the computation of the correlation coefficients becomes independent of E. However, from the practical point of view, an increase of the E-value also translates into a decrease of the time-lagged vectors' population, which in turn rarefies the related trajectories on the embedded manifolds and hinders the implementation of the algorithm. In light of this fact, the selection of the lowest possible E-value in all the examples that follow appears justified by the briefness of the diversities' time series, each constituted only by 26 records. In the next paragraph, the implementation of the dynamics model will open the possibility to confront the $E = 2$ case with more general scenarios, in order to detect possible improvements of the results.

Figures 32, 33 and 34 show the analysis of a heterogeneous bunch of countries equally distributed along the GDP ranking, emphasizing the patterns followed by the correlation coefficients as longer portions of the time series are implemented. In particular, the coefficients calculated at the generic point $L = n$ are related to time series truncated in correspondence of their n-th record. For each country, the analysis has considered the dynamical correlation between the GDP and, respectively, the export diversity, the import one and the total one. Each red curve refers to the estimation of one of the diversities' time series starting from the time-lagged vectors lying on M_{GDP}, while each blue curve refers to the opposite operation. The errors associated to the points of the curves reflect the precision of the estimation of the time series and will be clarified in Appendix A.

A first glance at the plots reveals a rather variegated set of scenarios in terms of both the final correlation levels and the trends followed by the curves, with a widespread dominance of the red ones. Then, observing the different diversity scenarios, no clear relation can be recognized between the export and the import columns of plots: some countries present higher correlation levels with the GDP if the export diversity is

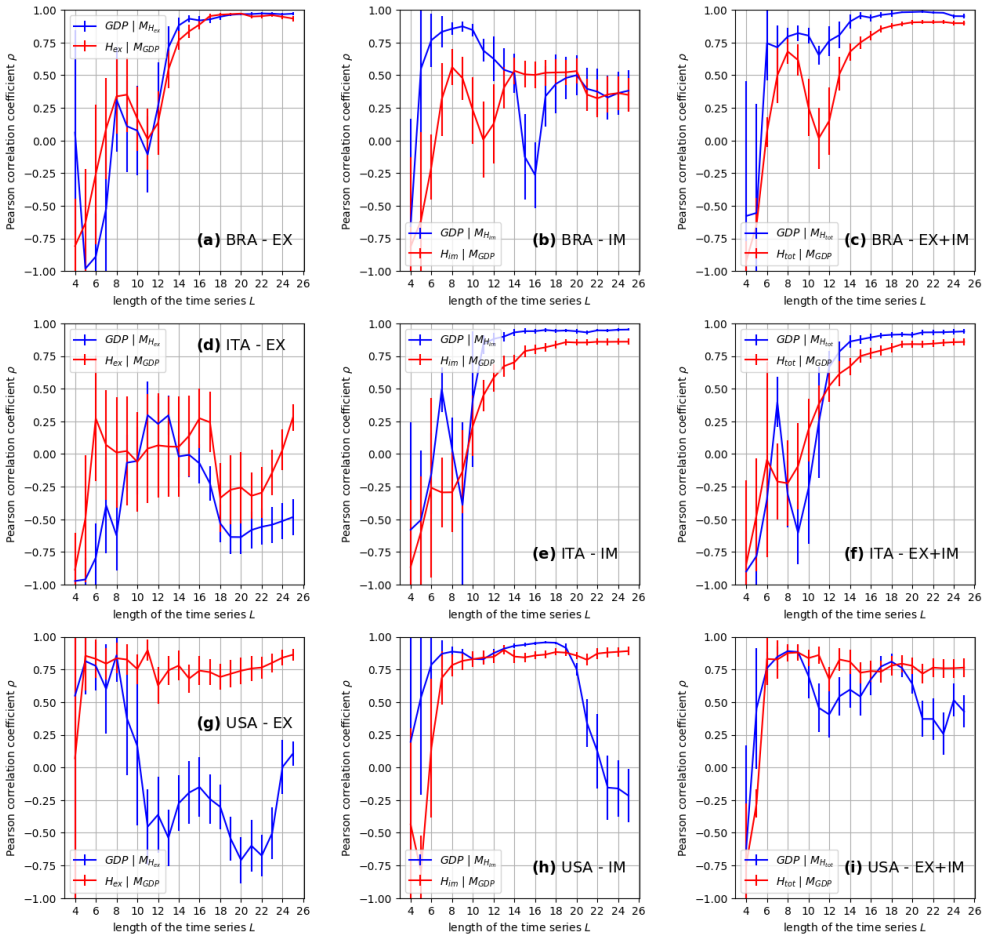

Figure 32: dynamical correlation between the GDP and the three diversities in the case of Brazil (BRA, **(a)**, **(b)** and **(c)**), Italy (ITA, **(d)**, **(e)** and **(f)**) and USA (USA, **(g)**, **(h)** and **(i)**). The blue curves describe the quality of the estimation of the GDP time series, while the red curves refer to the opposite operation.

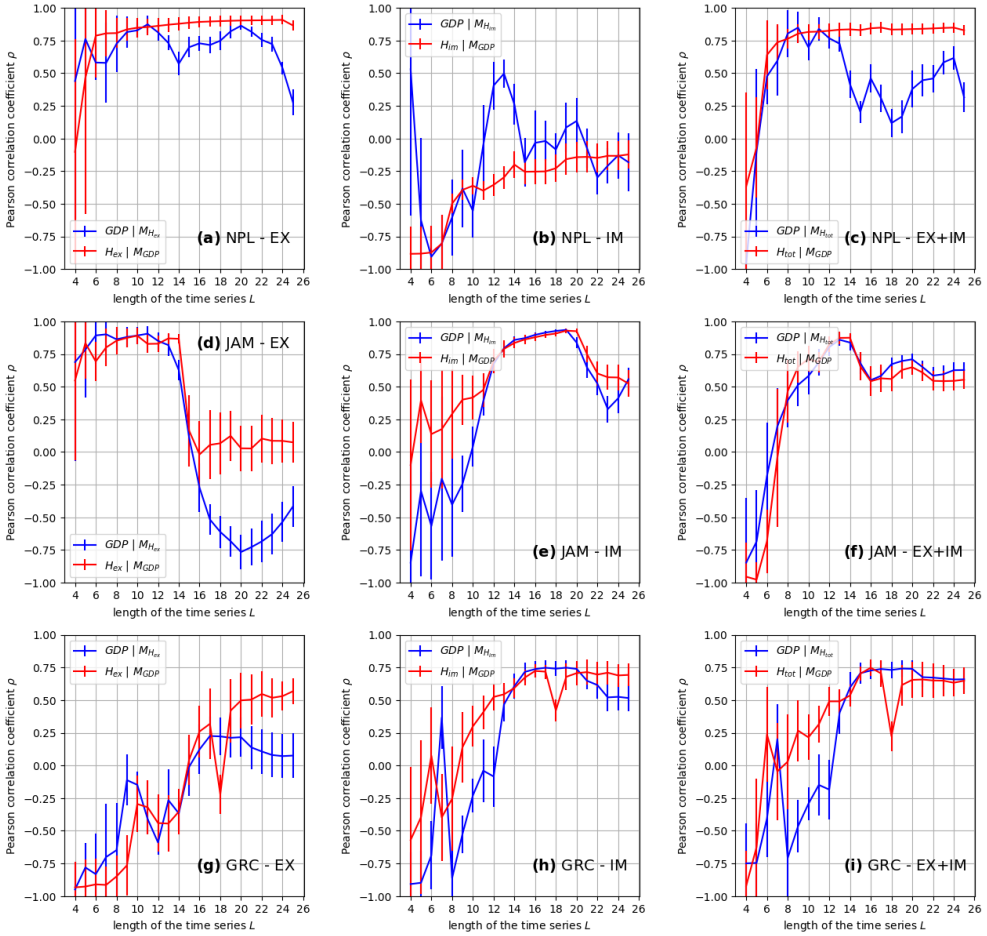

Figure 33: dynamical correlation between the GDP and the three diversities in the case of Nepal (NPL, **(a)**, **(b)** and **(c)**), Jamaica (JAM, **(d)**, **(e)** and **(f)**) and Greece (GRC, **(g)**, **(h)** and **(i)**). The blue curves describe the quality of the estimation of the GDP time series, while the red curves refer to the opposite operation.

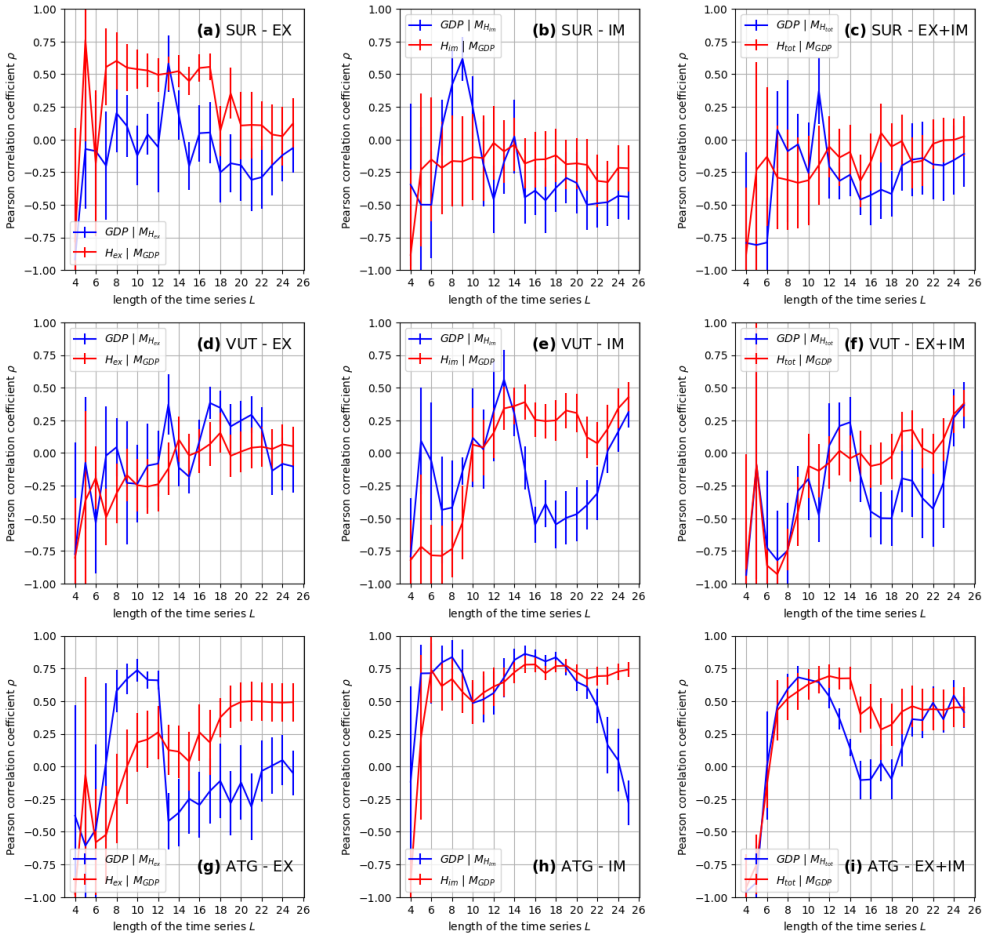

Figure 34: dynamical correlation between the GDP and the three diversities in the case of Suriname (SUR, **(a)**, **(b)** and **(c)**), Vanuatu (VUT, **(d)**, **(e)** and **(f)**) and Antigua and Barbuda (ATG, **(g)**, **(h)** and **(i)**). The blue curves describe the quality of the estimation of the GDP time series, while the red curves refer to the opposite operation.

considered, while some others are interested by the opposite property. Conversely, a remarkable feature shared by all the countries represented is the improvement provided by the implementation of the total diversity, which presents correlation levels that are always greater than or equal to both the export and the import ones. The nature of this improvement is quite varied. In many cases, the correlation levels related to the total diversity derive directly from either the export (Brazil, Nepal and Suriname) or the import (Italy, Jamaica and Vanuatu) ones and are not significantly hindered by the possible poor performance provided by the other measure; but in some others (United States, Greece and Antigua and Barbuda), the unification of the two single diversities proves to be able to generate a significant correlation boost between at least one of the two pairs of time series involved. All the countries are also joined by a high compatibility between the final correlation levels detected in the total diversity's scenarios, in contrast to the pronounced dissimilarity which can be often observed in the other cases.

Turning the focus on the distinction between the different countries involved, it is easy to note the vertical pattern followed by the total diversity's column of plots. The most developed countries' group, represented by Brazil, Italy and United States, exhibits clear-cut convergent trends towards generally higher final correlation levels, both between the GDP time series and the total diversity ones. In particular, the Italy's and the Brazil's scenarios clearly distinguish among the others for the achievement of almost perfect correlation levels in both the cases considered. On the other hand, less developed countries tend to display a poorer dynamical correlation between the GDP's and the total diversity's dynamics; in particular, the countries which occupy the lowest section of the GDP ranking present the lowest correlation levels and some of them, like Suriname, fail to exhibit any significant convergent trend as more empirical records are implemented into the analysis.

Remarkably, all the patterns just described are clearly perceivable also in Figure 35, which collects in separated rows of plots the average results related to different sets of countries. In particular, the first and the third rows show the average curves related to the countries which occupy, respectively, the first 10 positions and the last 10 positions inside the average GDP ranking extracted from the period 1995-2020 (whose compilation has excluded the countries which present holes inside the respective time series); both the top and the bottom section of this ranking are shown in Table 4. On the other hand, the intermediate row depicts the average curves extracted from the whole population of countries.

First of all, the extraction of the averages confirms the dominance of the red curves over the blue ones, especially in the top and in the average sections of the figure; moreover, an increase of the compatibility between the two kinds of curves can be clearly observed in all the diversity scenarios. Then, in confirmation of the horizontal trends observed in Figures 32, 33 and 34, the total diversity exhibits, on average, a stronger dynamical correlation with GDP with respect to the export and import measures alone; this property is particularly noticeable if the top section of the GDP ranking is considered, but slight improvements are also perceivable both in the average case and, to a lesser extent, in the bottom section. Furthermore, all the diversity scenarios confirm the clear separation between the top and the bottom sections of the GDP ranking in terms of the final correlation levels reached; in particular, the total

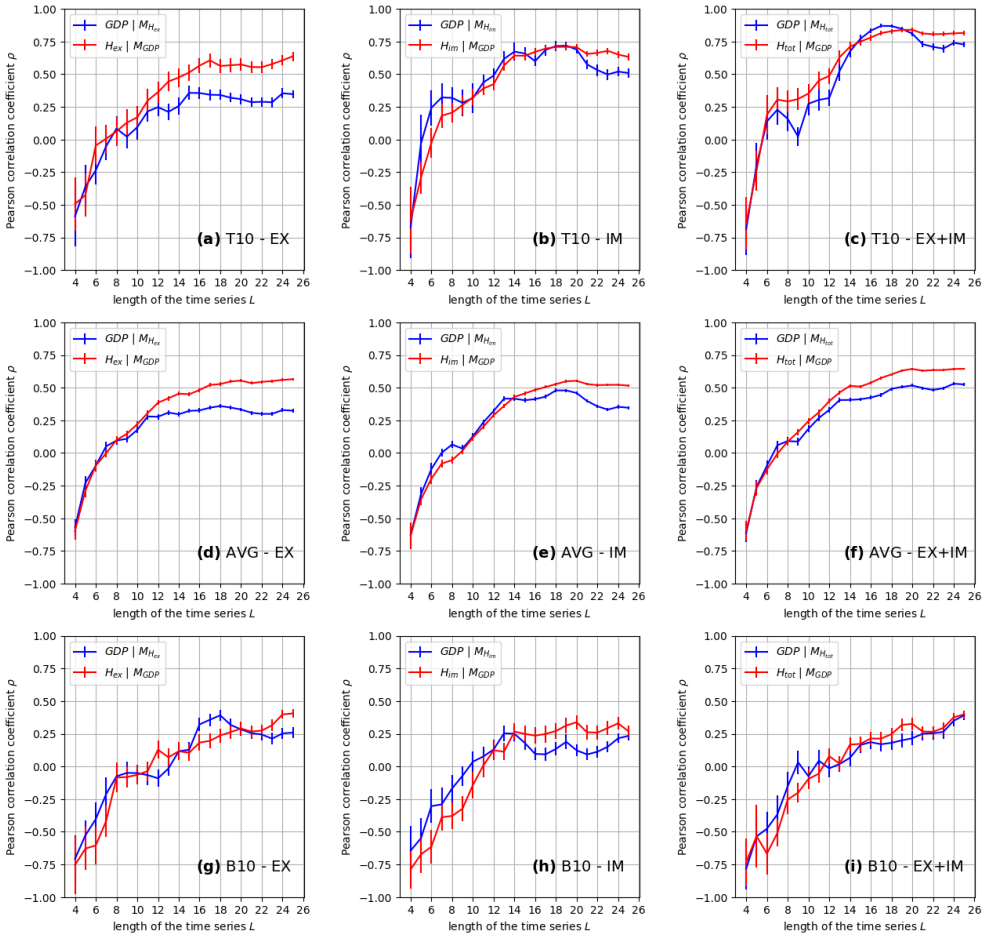

Figure 35: dynamical correlation between the GDP and the three diversities in the case of the top-10 section ((**a**), (**b**) and (**c**)) and the bottom-10 section ((**g**), (**h**) and (**i**)) of the average GDP ranking and in the average case ((**d**), (**e**) and (**f**)). The blue curves describe the quality of the estimation of the GDP time series, while the red curves refer to the opposite operation.

diversity's column of plots exhibits the most notable distinction in this sense.

The clear refinement of the curves and the increased compatibility between them allow to deem the results presented in Figure 35 as much more reliable than the single countries' ones, whose interpretation is always heavily affected by the briefness of the empirical time series. Firstly, the different extent of the correlation boosts generated by the unification of the single diversities, depending on the degree of development of the countries, can be traced back to the strength of the connection which binds the export branch with the import one. In particular, the clear-cut strengthening of the correlation levels in the case of the most developed countries can be ascribed to the strong interdependence between their two commercial branches. As already suggested by the analysis of the overlap between the export and the import rankings in Section 6.1, the inner composition of both the two baskets of a developed country is generally strongly influenced by the globally aggregated demand and supply law and thus reflects the needs of the global productive chain. In this situation, where the export and the import activities represent opposite sides of the same coin, it is reasonable to suppose them to evolve following alternating patterns: in particular, a local period of bullish growth of the export of a given market sector should be both anticipated and followed by analogous periods of growth of the related import and vice versa. In light of this fact, it seems natural to expect the unification of the export and the import branches to induce some sort of stabilization effect on the dynamics of the commercial activity, which can be deemed in turn as the origin of the correlation boost observed in Figure 35c. On the other hand, the analysis in section 6.1 has also suggested that, conversely, the import activity of an underdeveloped country generally reflects the needs of its society rather than the ones of the global market. Accordingly, the weak interdependence between the export and the import branches of such countries is unable to generate a stable dynamics by means of their unification, as the absence of a significant correlation boost in Figure 35i suggests.

In light of the discussion exposed at the end of the previous paragraph and of the evidence presented by Figure 35, the first important conclusion that can be drawn from the analysis is the general dominance of the potential causal influence exercised

Average GDP Ranking (1995-2020)			
Top-10 Section		**Bottom-10 Section**	
Average Rank	**Country**	**Average Rank**	**Country**
1.00	USA	158.77	Saint Kitts and Nevis
2.42	Japan	159.46	Saint Vincent and the Grenadines
3.54	Germany	161.23	Samoa
4.08	China	161.38	Vanuatu
4.81	United Kingdom	162.00	Dominica
5.50	France	164.27	Tonga
7.35	Italy	164.62	Federated State of Micronesia
9.27	Brazil	166.12	Marshall Islands
9.54	Canada	166.88	Kiribati
10.92	India	168.00	Tuvalu

Table 4: Top-10 and bottom-10 sections of the average GDP ranking extracted from the period 1995-2020.

by the diversities on the dynamics of GDP over the one that regulates the opposite relations: this feature could be easily explained by observing that relevant changes in the diversification of the commercial activity can often exercise a direct influence on the wealth of a country (for example, in the context of the opening of new market sectors), while possible surges in the evolution of the latter often do not generate direct repercussions on the distribution of the traded products.

Then, another important conclusion deducible, in particular, from Figure 35c is the statement of the potential presence of a bidirectional causal relation between the total diversity and the GDP of the most developed countries. On one hand, the potential causal influence exercised by the total diversity on the dynamics of GDP could be justified by the stronger connection existing between the commercial activity of these countries and the needs of the global productive chain, which generates more direct effects on their societies in the form of job and investments opportunities; on the other hand, the opposite casual relation could be attributed to the investment policies typically adopted by the same countries, which generally show the tendency to reinvest a considerable fraction of their income into their commercial activity. Conversely, the poor correlation levels detected in the third rows of plots of Figure 35 lead to conclude that, most likely, similar causal relations cannot be stated in the case of the less developed countries. A simple explanation of this situation can be obtained by considering that these countries typically lack an internal organization able to implement adequate investment policies and their baskets are usually dominated by a few market sectors: for this last reason, a negative event like the considerable drop of either the export or the import of a key sector results in an increase of the diversity of the related basket and generates ambiguous correlation signals between the diversification of the commercial activity and the wealth.

At this point, it is intriguing to check the soundness of the results presented in Figure 35 by analyzing the dynamical correlation between the diversities and an economic indicator of different kind. For example, an interesting test consists in the implementation of an intensive indicator like GDPpc, in order to verify the effects of the introduction of factors (in this case, the total population) which do not manifest any clear relationship with the economic activity of a country. To this purpose, Figure 36 presents a scheme of plots which is completely analogous to the one of Figure 35: the first and the third rows represent again the top-10 and the bottom-10 sections of the average GDP ranking, while the intermediate row presents the average curves extracted from the whole population of countries. The comparison between Figures 35 and 36 reveals, first of all, an extreme similarity between the rows devoted to the group of most developed countries. This feature, perceivable not only in the magnitude of the final correlation levels but also in the global structure of each curve, is linked to the high degree of stability which characterizes the dynamics of the populations of these countries: the obvious consequence of this property, indeed, is the substantial equivalence, in terms of local fluctuations, between the dynamics of GDP and GDPpc. On the other hand, the population proves to represent a much more destabilizing factor when less developed countries are considered, as can be deduced by the observation of the second and the third rows of Figure 36. Both the scenarios, indeed, present much weaker correlation levels with respect to the ones related to GDP; moreover, the unification of the export and the import diversities generates a much softer improve-

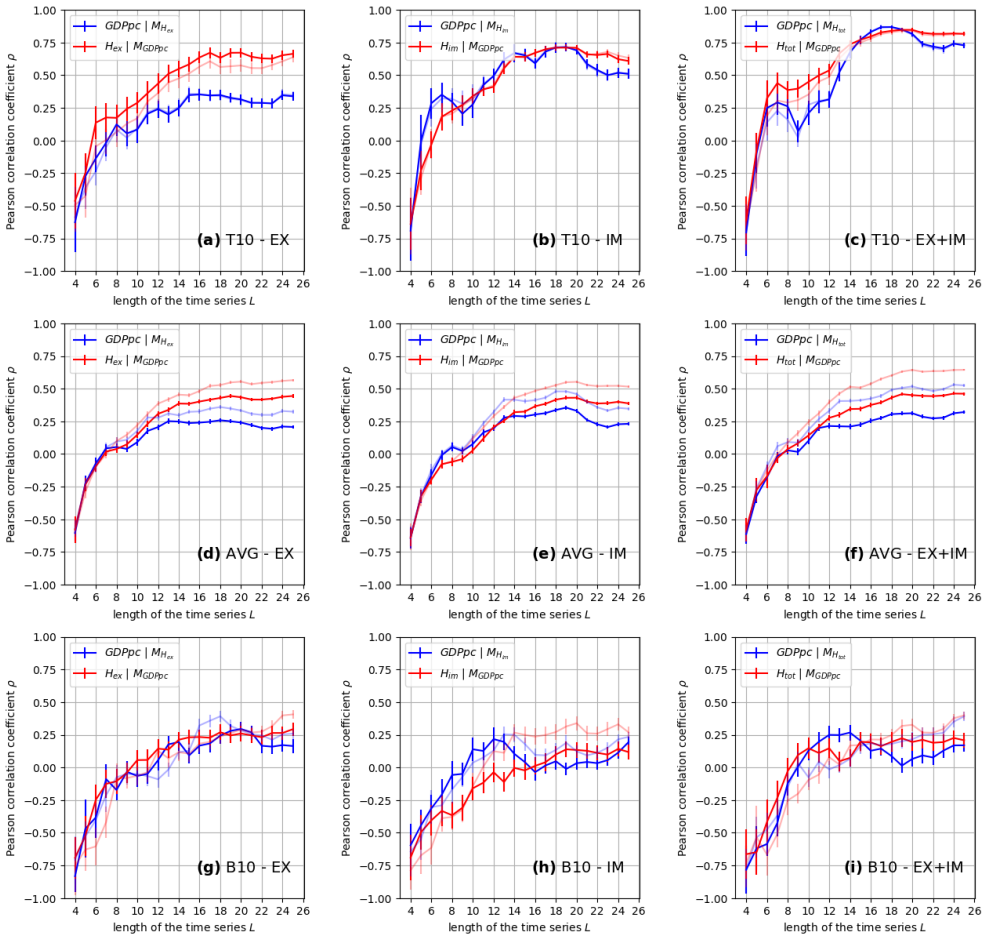

Figure 36: dynamical correlation between the GDPpc and the three diversities in the case of the top-10 section ((**a**), (**b**) and (**c**)) and the bottom-10 section ((**g**), (**h**) and (**i**)) of the average GDP ranking and in the average case ((**d**), (**e**) and (**f**)). The blue curves describe the quality of the estimation of the GDPpc time series, while the red curves refer to the opposite operation. The curves related to the GDP are depicted on the background.

ment of the correlation levels in the average case and no perceivable improvement in the bottom section. In line with the interpretation provided before, the poorness of these results can be attributed to the relatively small and unstable population which is typical of an average country and, even more so, of an underdeveloped one. In the end, the distinction between Figure 35 and Figure 36 allows to state the soundness of the dynamical correlation between the diversities and the GDP of a country and to confirm the extensive character of the former.

6.2.3 Synthetic dynamical correlation

Throughout the previous paragraph, the briefness of the empirical time series has been described as the most relevant obstacle to the interpretation of the correlation curves. In light of the positive results obtained during the validation of the dynamic model in Section 5.1, the proposal of the generation of detailed, synthetic diversity time series, with the purpose of reproducing and, possibly, expanding the empirical ones, seems quite intriguing and represents the main subject of this paragraph. The presentation of the results, however, requires an exhaustive explanation of all the limitations tied to the application of the dynamic model.

The main issue in this context concerns directly the reproduction of the diversities' time series in a single country's scenario. On one hand, the results presented in Figure 11 have proved the ability of the transfer mechanism embodied in the model to produce, on average, a clear-cut convergence of the initial disposition of the products in a basket towards the final empirical ranking (see Eq. 2.1). Remarkably, as can be noticed in all the three cases portrayed in the figure, the typical smallness of the characteristic times (always of the order of a ten of years) allows a significant approach of the average Spearman's coefficient towards the asymptotic value already after 25 years; moreover, crucially, each plot also shows the absence of relevant fluctuations in the evolution of the single coefficients during the same time interval. As a result, the combination of these features with the correct assessment of the final monetary value of a basket (by means of a proper calibration of the drift parameter $\overline{\mu}^c$) suggests the potential ability of the model to provide a plausible estimation of the final bare diversity of the basket itself as the outcome of a single 25 years simulation, at least in the case of a very developed country. However, as can be noted even from Eq. 3.1 alone, the model does not embody any constraint on the intermediate steps of the evolution of the trades and, thus, it does not address in any sense the reproduction of their whole dynamics; as a consequence, exception made for the first and the last records, the faithful reproduction of the diversity's empirical time series in their totality is clearly out of range. Nevertheless, this condition does not negate on its own the possibility of the reproduction of the average correlation curves related to a certain group of countries.

Then, a second limitation concerns the subset of countries which are eligible for the analysis. As explained in Chapter 2, every hole in the aggregated database represents an unmanageable singularity for the implementation of the model and must be discarded, along with the whole time series which contains it. Usually, the exclusion of even a considerable number of products from the analysis does not significantly hinder the calibration of the dynamical parameters; conversely, in the context of the

reproduction of the empirical diversity's records, the high sensitivity of the Shannon's entropy function to the length of the probability vectors implies the need of limiting as much as possible the expulsion of products from the baskets. For this reason, the analysis presented in this paragraph has considered only the countries which occupy the first 10 positions inside the average GDP ranking (see Table 4): as the intuition suggests, the baskets related to these countries exhibit strong stability in the evolution of their trades and are interested by very limited numbers of holes.

Finally, a couple of issues interest the production of finer time series, characterized, for example, by a semi-annual frequency. In the diversity case, the main issue concerns the refinement of the bare records simulated by the model. While the annual records of each time series can be refined by simply implementing them as initial conditions of the iterative algorithm (see Eqs. 4.4 and 6.4) in place of their empirical counterparts, the absence of a complete network (empirical or synthetic) of diversity and ubiquity values prevents a rigorous refinement of the semi-annual ones. In addition, the production of semi-annual GDP time series is severely hindered by the absence of detailed databases, which interests almost every country in the network. For these reasons, it is necessary to implement adequate estimation procedures of the missing records. In order to preserve as much as possible the dynamics of the bare diversity time series, each of the bare semi-annual records has been multiplied by the average of the correcting factors related to the surrounding annual ones, which take the general form $H_c/H_c^{(0)}$ (see Eq. 4.11). However, since the interest of the analysis lies in the local dynamics of the time series rather than on the magnitude of their records, the refinement of the bare series, as long as it does not alter their evolution, does not play a relevant role. On the other hand, each GDP semi-annual record has been estimated through the average of the surrounding annual ones.

A short paragraph must be devoted, now, to a brief description of the application of the dynamic model to the import activity of the countries. Given the complete equivalence, in terms of the adopted aggregation procedures, of the respective final databases (see Section 6.1), the only difference between the import and the export cases concerns the interpretation of the parameters of the model. In this regard, the discussion presented in Sections 6.1 and 6.2 has already provided some useful concepts. Generally, in relation to their export equivalents, all the import parameters are more strongly influenced by factors which are external to the country they refer to. First of all, the drift parameter $\overline{\mu}^c$ represents the average growth rate of the internal demand of products, which can be attributed both to the increase of the actual necessity of goods (due, for example, to the growth of the population) and, in the case of a rapidly emerging economy, to the increase of the degree of integration into the global productive system. Turning to the stochastic component of the evolution, σ^c and τ^c describe, respectively, the extent of the fluctuations of the aforementioned demand and the duration of the related periods of growth and stagnation; on the other hand, σ_0^c quantifies once again the effects of the turbulence of the external market on the evolution of the baskets and presents, accordingly, an inverse proportionality with the degree of integration into the aforementioned global system. Lastly, the definition of the transfer parameter G^c provides the most evident interpretative distinction between the export and the import cases. The transfer of resources inside an import basket, indeed, is heavily influenced by the passive branch of collective dynamics ascribable

to the internal needs of the society of the respective country; as a consequence, G^c cannot be entirely related to the investment policies implemented by a government and loses a consistent part of the prescriptive nature which is typical of its export counterpart, especially in the case of an underdeveloped country. Nevertheless, the transfer of resources can be still actively modified, according to the needs of the global market, in order to exploit the correlation between the external demands of different products and to enhance the contribution to the growth quantified by the parameter h_T^c.

Table 5 shows the results of the calibration of the selected import baskets and compares them with their export equivalents, in order to outline possible analogies and differences between the two cases. In place of the bare ones, the table considers the same composite parameters which have been presented in Figure 24. Firstly, a general feature that can be easily observed is the substantial compatibility between the export and the import ranges of values: taking the export ones as a reference, no significant outliers can be observed in any of the import sections. In addition, in many cases the export and the import branches present a clear overlap. This feature is especially evident along the σ_{tot}^c column, where, with the only exceptions of China and France, the results indicate that the export and the import baskets of a very developed country generally share very similar degrees of volatility, both of internal and external

Export calibrated parameters				
Country	$\langle \log\left(J_{pp'}^c\right)\rangle$	σ_{tot}^c	$\overline{\mu}^c$	h_T^c
USA	-5.220 ± 0.007	0.49 ± 0.03	-0.0040 ± 0.0003	0.0012 ± 0.0006
Japan	-6.66 ± 0.03	0.59 ± 0.02	-0.0082 ± 0.0005	0.0024 ± 0.0005
Germany	-5.018 ± 0.006	0.35 ± 0.06	0.0109 ± 0.0008	0.0009 ± 0.0004
China	-5.447 ± 0.007	0.74 ± 0.05	0.0651 ± 0.0006	0.0056 ± 0.0008
United Kingdom	-5.181 ± 0.006	0.52 ± 0.02	-0.0190 ± 0.0005	-0.001 ± 0.001
France	-4.884 ± 0.005	0.63 ± 0.05	-0.0145 ± 0.0007	-0.0014 ± 0.0006
Italy	-5.60 ± 0.01	0.52 ± 0.05	0.0010 ± 0.0004	0.0039 ± 0.0007
Brazil	-7.418 ± 0.009	0.98 ± 0.04	0.0078 ± 0.0004	0.0005 ± 0.0005
Canada	-6.65 ± 0.02	0.66 ± 0.05	-0.0026 ± 0.0005	0.0041 ± 0.0003
India	-5.873 ± 0.005	0.94 ± 0.02	0.0620 ± 0.0009	0.0009 ± 0.0005
Import calibrated parameters				
Country	$\langle \log\left(J_{pp'}^c\right)\rangle$	σ_{tot}^c	$\overline{\mu}^c$	h_T^c
USA	-4.815 ± 0.005	0.43 ± 0.06	0.0151 ± 0.0005	0.0007 ± 0.0008
Japan	-5.085 ± 0.005	0.52 ± 0.04	-0.0045 ± 0.0004	0.0012 ± 0.0009
Germany	-7.3 ± 0.4	0.37 ± 0.04	0.0015 ± 0.0004	-0.002 ± 0.001
China	-6.52 ± 0.03	0.95 ± 0.05	0.0648 ± 0.0008	0.001 ± 0.002
United Kingdom	-5.85 ± 0.01	0.52 ± 0.04	-0.0052 ± 0.0006	0.0009 ± 0.0007
France	-5.854 ± 0.009	0.35 ± 0.05	-0.0048 ± 0.0005	-0.0005 ± 0.0008
Italy	-6.82 ± 0.02	0.46 ± 0.06	0.0026 ± 0.0008	0.0009 ± 0.0007
Brazil	-7.13 ± 0.01	0.92 ± 0.03	0.0132 ± 0.0004	-0.0006 ± 0.0008
Canada	-7.8 ± 0.2	0.47 ± 0.03	0.0100 ± 0.0005	0.0002 ± 0.0006
India	-7.24 ± 0.09	1.03 ± 0.03	0.0757 ± 0.0009	0.0019 ± 0.0005

Table 5: Calibrated parameters of both the export and the import baskets related to the top-10 section of the average GDP ranking.

nature. A diffuse similarity can be observed also across the $\overline{\mu}^c$ column. Here, the notable cases of China and India, probably the countries which have been interested by the most impressive economic growth during the last decades, present an excellent compatibility between their values, to be related both to the massive growth of their population and the strongly increasing degree of integration into the global productive system. In the case of more stable countries, very often the stronger influence of factors independent of the pure commercial activity results in a more variegated population of values. Then, although a clear relation cannot be identified, in every case both the export and the import baskets are characterized by very low values of the parameter h_T^c, denoting the general irrelevance of the transfer mechanism in the computation of the overall growth. Finally, the formation of a clear pattern along the $\log\langle J_{pp'}^c \rangle$ column is prevented by the presence of the branch of dynamics which is ascribable to the internal demand of needed goods, highly variable depending on the country.

At this point, the time series generated by the dynamic model and refined through the procedure described above can be finally presented. The bare version of each of them is the outcome of a single simulation of the model: this choice allows to obtain a dynamics as varied as possible and avoids the excessive flattening which would result from the implementation of the average of multiple simulations. In Figure 37, the synthetic curves and their empirical counterparts are disposed, respectively, along the second and the first row of plots, while the columns distinguish the three different diversity scenarios. As expected, the synthetic curves generally exhibit a much more definite smoothness and are unable to replicate the whole variety of local fluctuations typical of the empirical ones. However, in confirmation of the effectiveness of the transfer mechanism incorporated in the model, the quality of the reproduction of the general trends followed by the empirical time series seems quite satisfactory: for example, the standout trends of the Brazil's export diversity and the China's import one, both sharply negative, appear to be optimally reproduced. Moreover, the observed overall compatibility between the empirical and the synthetic records confirms the validity of the adopted refinement procedure.

By implementing the time series depicted in Figure 37, Figure 38 finally presents the synthetic reproductions of the average correlation curves related to the top-10 section of the GDP ranking. Each curve describes the correlation between the dynamics of the empirical GDP time series and the one of the synthetic time series related to one of the three diversity measures. Adopting the same theoretical setting seen in the previous paragraph, the manifolds resulting from Takens's theorem have been embedded into a minimal embedding space ($E = 2$). While the export and the import synthetic correlation levels slightly overcome their empirical counterparts, the ones related to the total diversity present a strong compatibility with the ones shown in Figure 35c. Furthermore, despite the raising of the export and the import curves, the synthetic scenario, along the lines of the real one, still exhibits a clear improvement of the correlation levels generated by the unification of the single diversities. In conclusion, the results presented in Figure 38 allow to state the ability of the model to reproduce, on average, the dynamical correlation between the diversities and the GDP of the most developed countries. This important property, in turn, leads to a couple of additional considerations. On one hand, the results provided by the model confirm once more the soundness of the empirical ones and allow to deem the minimal state

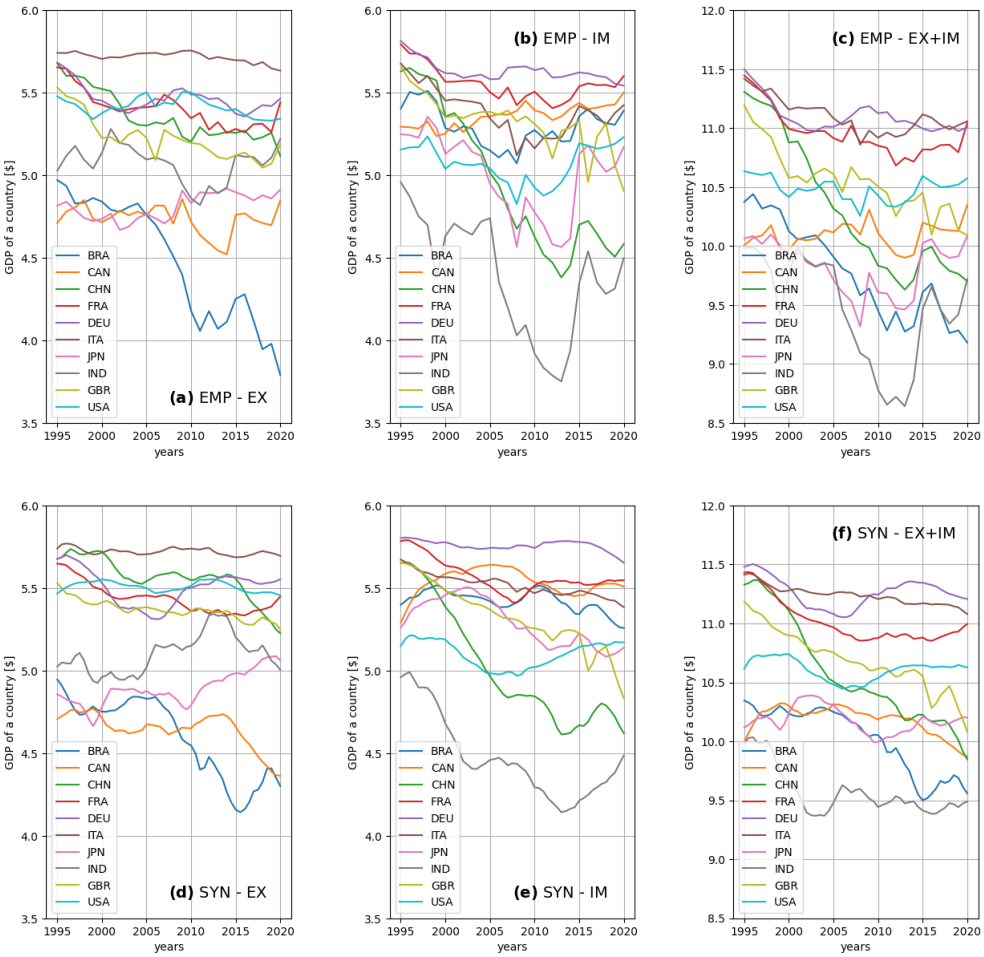

Figure 37: evolution across the period 1995-2020 of the empirical time series ((**a**), (**b**) and (**c**)) and of the refined synthetic ones ((**d**), (**e**) and (**f**)) related to the three diversities.

space scenario as an excellent approximation to be employed in many practical applications, especially in condition of scarceness of empirical records; moreover, the same results lay the foundation for the employment of the model as a reliable forecasting tool, with the purpose of generating meaningful predictions on the future wealth of the most developed countries entirely based on the dynamics of their global commercial activity.

The high accuracy of the reproduction of the empirical correlation curves also justifies the expansion of the analysis through the implementation of more detailed synthetic time series, with the purpose of comparing the minimal state space scenario with more general ones. As explained in the previous paragraph, increasing the E-value

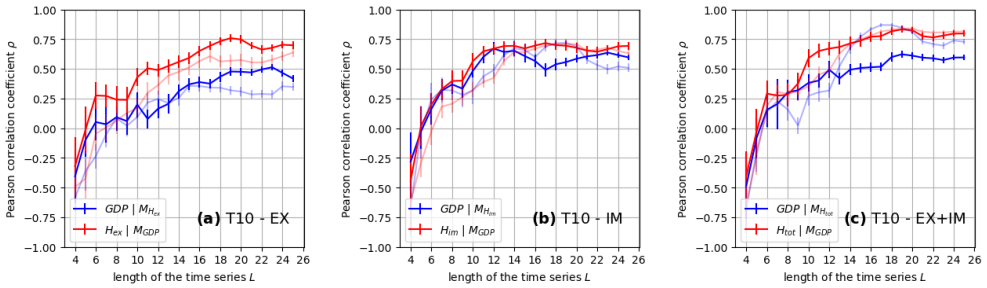

Figure 38: simulated dynamical correlation between the GDP and the three diversities in the case of the top-10 section of the average GDP ranking. The blue curves describe the quality of the estimation of the GDP time series, while the red curves refer to the opposite operation. The empirical curves are depicted on the background.

means tightening the dynamics of the variables involved and improving the reliability of the results. However, the consequent increase of the length of the time-lagged vectors implies a reduction of the fraction of the time series which is available for the analysis and can represent, in condition of scarceness of empirical records, a relevant obstacle to the application of the algorithm. In this context, the implementation of more detailed time series allows to decrease the relative importance of such reduction and, then, to safely increase the E-value. Adopting the procedures described previously, refined time series with semi-annual frequency for both the GDP and the single diversities have been extracted from the same set of simulations employed before. Finally, Figure 39 presents the respective correlation curves, devoting each row of plots to a different E-value. First of all, the comparison between the first row of the figure and Figure 38 reveals a clear improvement of the correlation levels and an increase of the smoothness of the trends followed by the curves in all the diversity scenarios: both these features, if combined with the verified reliability of the model outcomes, increase the likelihood of the bidirectional causal relation potentially existing between all the three diversities and the GDP of the most developed countries, allowing to strengthen the interpretation of the results provided at the end of the previous paragraph. On the other hand, the subsequent rows of the figure allow to perceive the progressively more accentuated improvement of all the correlation levels provided by the increase of the E-values, as a direct consequence of the growth of the embedding space and of the strengthening of the connection between the dynamics of the variables. Lastly, the soundness of the unification of the single diversities is ultimately confirmed, as a consequence of the presence of a clear correlation boost in all the scenarios described by Figure 39.

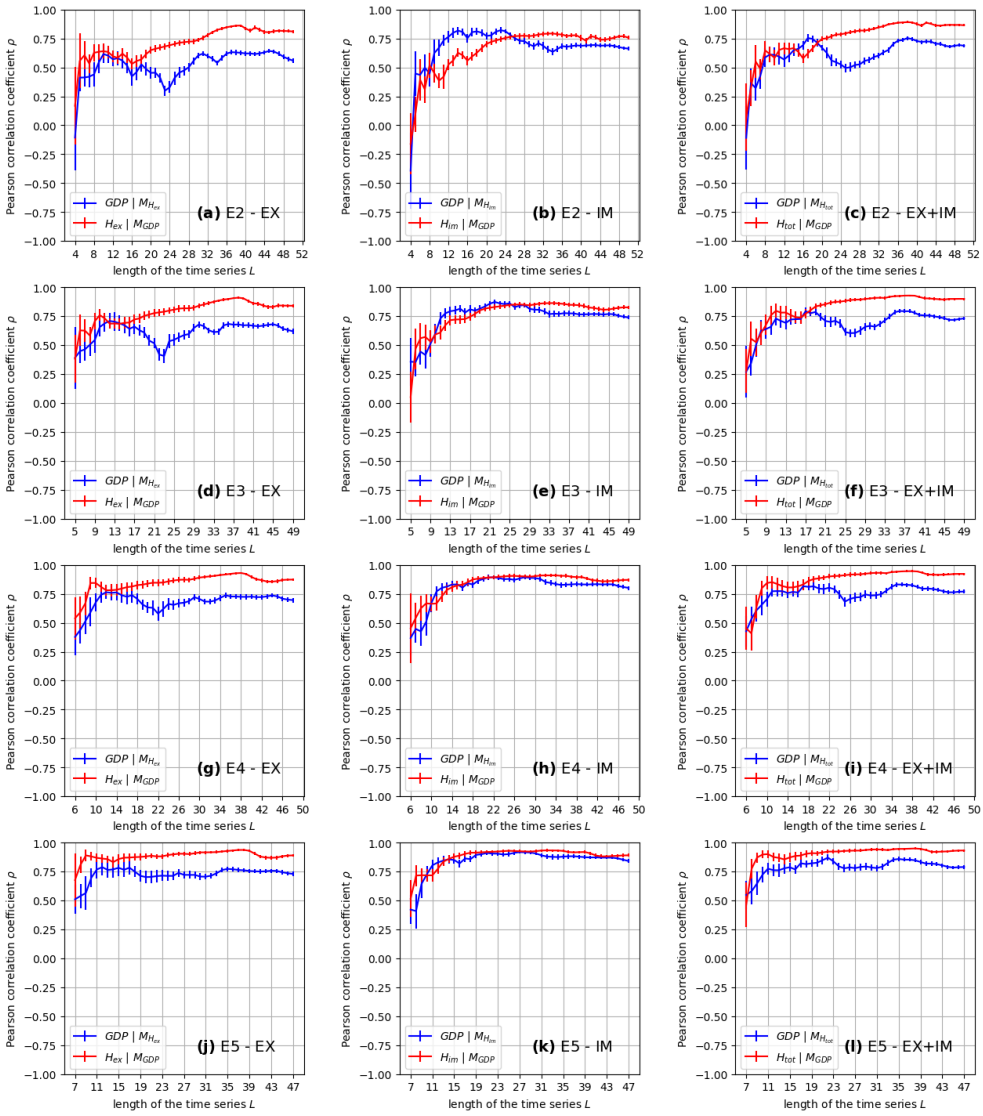

Figure 39: detailed simulation with semi-annual frequency of the dynamical correlation between the GDP and the three diversities in the case of the top-10 section of the average GDP ranking, for different E-values. The blue curves describe the quality of the estimation of the GDP time series, while the red curves refer to the opposite operation.

7 Conclusions

The analysis of BACI through the dynamic model and the complexity measures introduced in the articles by Stella, Caraglio and Teza [4, 14, 15] has provided substantial contributions to the econophysics field, leading to a deeper understanding of economic complexity. In particular, the cooperation between the two analytic methods has brought to an enrichment of the characterization of the bipartite network employed for the description of the global trading system and to a considerable extension of its scope, overcoming at the same time the conceptual and technical issues affecting the approaches proposed by Hidalgo and Hausmann in [16] and by Pietronero in [3,17,18].

Inspired from the beginning by a minimalist conception, the ensemble of parameters governing the equations of the dynamic model offers a concise but still thorough representation of the evolution of the national baskets and, indirectly, of the factors influencing the commercial activity of the countries. In the first place, the independent exponential growth rate which interests, on average, each trade in a given basket is effectively described by the drift coefficient μ: in particular, the latter embodies a time dependent component representing the inflationary growth of prices and a constant one consisting in the drift parameter $\overline{\mu}^c$, accounting for the financial investments in the productive sector in the export scenario and for the growth of the internal demand in the import one. Then, the noise parameters σ and τ^c work in coupling in order to describe the influence of random low-scale events on the volatility of the evolution of the basket, representing respectively the extent of the fluctuations and the duration of related periods of persisting growth or recession; on the other hand, the additional noise parameter σ_0^c, introduced in [14] with the purpose of improving the calibration procedure, quantifies the dynamical effects generated on the basket by the global productive system. Lastly, the extent of the transfer mechanism embodied in the model is tuned by the parameter G^c, which assesses the strength of the mutual dependence between different productive chains or the internal demands of similar products. Furthermore, a remarkable peculiarity of the model, in contrast to other methods, consists in the incorporation of the time correlation between the historical trading records into both the transfer mechanism and the stochastic component of each equation, with the purpose of reaching an optimal matching degree with the empirical evolution of the trades. Another advantage stemming from the minimalist spirit of the model is the simplicity of the calibration procedure, whose phases can be progressively reached through simple reformulations of the equations of the model. In this regard, moreover, each phase of the procedure has provided unambiguous and reliable parameters for all the countries considered in the analysis.

Following the calibration procedure, the dynamic model has demonstrated a clear ability to supply faithful reproductions of the empirical evolution of the baskets, both in the export and in the import scenarios (explicitly shown in the case of United States, China and Russia). In particular, a quick comparison between the empirical plots and their respective synthetic counterparts is sufficient to perceive a pronounced similarity between the overall trends and the degrees of volatility. Furthermore, on the quantitative level, the reproductions of the model in the export case have shown an

optimal compatibility with the empirical scenarios in terms of the average growth rates, the time correlation between the trades and the final rankings of the products. This last feature, in particular, has been explicitly demonstrated in the case of United States, China and Russia through the study of the evolution of the Spearman's correlation coefficients over time; in addition, the same study has also provided an estimation of the period of time required for performing substantial structural alterations of the productive system of a country, which, in the three cases considered, turns out to be of the order of a ten of years. Visually, the convergence of the initial distributions of the trades towards the final rankings can be deduced from the strengthening of the rainbow effects over time and from the correct reproductions of the most increasing trends.

In the evaluation of the average growth rate of a basket, the dynamic model proposes an assessment of the contribution provided by the branch of collective dynamics of the trades, which stems from the interplay between the random fluctuations and the transfer mechanism. As shown by the counterfactual analysis of the export growth prospects, the relative extent of such contribution (with respect to the drift parameter $\overline{\mu}^c$ and the average inflationary term) presents a variegated dependence on the transfer parameter G^c, depending on the historical stability of the national economy considered. In particular, economies characterized by a long-term consolidated structure present a marginal contribution from the transfer of resources in the real scenario and a severe weakening for high values of G^c; on the other hand, in the same region of values, economies undergoing substantial structural changes exhibit from moderate to tremendous enhancements originating from the same transfer. The determination of the optimal rate of transfer of resources between different market sectors represents an example of the exploration-exploitation trade-off problems that can be arguably addressed through the implementation of the dynamic model.

Turning to the complementary side of the analysis, the implementation of the *diversity* and *ubiquity* complexity measures provides a further confirmation of the flexibility of the Shannon's entropy function and of its usefulness in the economic field. The bare versions of the measures, formally equivalent to the function, have undergone a refinement process consisting in the application of an iterative algorithm, with the purpose of implementing the intertwined nature of the global productive system into the diversification concept. From the technical point of view, the smoothness of the mathematical characterization of the algorithm as a real contraction has allowed to lay sound theoretical foundations in support of its convergent trend towards a unique fixed point, which has been successfully verified both in the empirical framework and in a sizeable sample of randomly selected synthetic scenarios.

Subsequently, the refined complexity measures have been employed in a novel characterization of the overall diversification of the global productive system, from both the export and the import perspectives. In the case of the countries network, the implementation of the export diversity (inspired by a preliminary analysis concerning the medians of the exports distributions) has realized a clear distinction between the group of most developed countries, generally characterized by a highly diversified export activity, and the emblematic poverty trap area, peculiarly affected by the domination of

few market sectors. This feature allows to state the reliability of the export diversity as an economic efficiency estimator; conversely, due to the generalized homogeneity of the import activity across the network, the same conclusion does not hold in the case of the import diversity. Furthermore, the analysis of the relationship between market sectors placed at different aggregation levels has allowed to subdivide the export diversity into inter-sectorial and intra-sectorial contributions and has suggested a possible employment of the latter as an indicator of the degree of internal organization of an economy, which supplies an additional criterion of distinction between developed and underdeveloped countries.

On the other hand, the implementation of the export and the import ubiquities has led to the definition of four different market regions, which, due to the high overlap degree detected between the two economic scenarios, provide a measure of the overall relevance of the products inside the network. Remarkably, the comparison between this representation of the global market and the composition of the national export baskets has also allowed to determine the conceptual direction of the investments made by the countries: while the underdeveloped ones are forced by the weakness of their economy to concentrate their productive efforts on the most demanded products, the abundance of the baskets exported by the most developed ones highlights their propensity to export all the products on the market, including the most marginal ones.

Afterwards, the conjunction of the two methods of analysis with the macro-economic narrative has resulted in a novel dynamical characterization of the countries network, which represents the pinnacle of the research presented in the articles by Stella, Caraglio and Teza [4, 14, 15]. Specifically, such result takes the concrete form of a particular two-dimensional benchmark, defined from the export diversity and the intensive economic indicator GDPpc. Due to the lack of correlation between the quantities on the axes, the distribution of the countries on the benchmark has allowed the subdivision the network into four different macro-regions, related to just as many combinations of degrees of general wealth and economic stability. Furthermore, in addition to the static characterization of the network in a given year, the study of the evolution of the countries across the benchmark during a certain period of interest has provided an assessment of the degree of volatility of each region: in this regard, the paths followed by the group of the most developed countries tend to present a reduced length and a smooth character, whereas the ones appearing in the other regions reflect much more turbulent dynamics. Finally, the representation provided by the benchmark has been enriched through the implementation of an additional dimension, defined from a single parameter related to the dynamics of the export activity: in this framework, the comparison between the distribution of the countries on the benchmark and the calibrated values of the parameters has supplied further effective distinctions between the macro-regions, based on multiple factors related to the export activity.

Lastly, the final part of this work has been devoted to an original analysis of the causal connection between the commercial activity and the global wealth of the countries. Following the same heuristic approach usually adopted in this context, such task

has been pursued through the evaluation of the dynamical correlation between various couples of complexity measures and economic indicators, chosen as representatives of the quantities of interest. From the practical point of view, the analysis has been conducted through the application of the convergent cross-mapping method method by Sugihara [30], whose theoretical background stems from embedding theory and fractal geometry.

Before addressing the main subject of the study, a preliminary analysis has unveiled interesting analogies and differences between the export and the import activities of the countries, providing useful insights on the economic mechanisms which subtend their evolution. In this regard, in particular, the unification of the export and the import diversities has resulted in the definition of a novel, comprehensive diversification measure, which works as a clearer estimator of the economic efficiency across the countries network.

Subsequently, the convergent cross-mapping method has been applied to the empirical time series of GDP and all the three diversities in several relevant case studies. In particular, the analysis has first concerned meaningful single-country cases and then a set of collective scenarios, related to the whole network of countries and to opposite sections of the GDP ranking. In the end, the most important conclusion that has been drawn from the analysis is the statement of the potential presence, in the case of a very developed country, of a significant mutual causal relation between the diversification of its commercial activity and its total income; conversely, mainly due to the instability of the respective economies, the opposite conclusion probably holds in the case of an underdeveloped one. Another interesting result consists in the detection, in all the collective scenarios analyzed, of a stronger dynamical correlation between the GDP and the total diversity, likely attributable to the stabilization effect generated by the unification of the two branches of commercial activity. Remarkably, the soundness of all these results and the extensive character of the complexity measures have been confirmed through the application of the method to the GDPpc empirical time series, which has provided poorer results due to the implementation into the analysis of a factor external to the commercial activity of the countries (namely, their total population).

Finally, the realization of a detailed, collective scenario related to top section of the GDP ranking by means of the application of the dynamic model has supplied a further confirmation of the reliability of the empirical analysis and has allowed to appreciate, from a different perspective, the ability of the model itself to faithfully reproduce the evolution of the commercial activity of the countries.

As a conclusive remark, the generality of the development of the analytic methods presented in this work opens the way to their application in several contexts, leading to an extension of the scope of the trading bipartite network. In this regard, an instance of interesting application of the dynamic model could consist in its incorporation into an in-depth study of the current degree of stability of a national economy, with the purpose of determining the rate of transfer of resources related to the optimal enhancement of its export activity. On the other hand, the flexible character of the complexity measures, resulting from the implementation of the Shannon's entropy function in

their construction, certainly supports their candidature for the quantification of the diversification of other systems related to the economic field (like the production chain of patents) or even detached from it.

Furthermore, the last part of the analysis has explored the dynamical properties of the complexity measures and their connection with the dynamic model, unveiling the predictive potential of both the methods. In the first place, for example, the proved ability of the model in the assessment of the dynamical correlation leads to the possibility of its employment as a tool for forecasting the future income of the most developed countries. Moreover, the detection of a strong dynamical correlation between the diversification of the baskets and the total income, possibly subtending an effective causal relation between them, suggests to employ the diversity measures in the characterization of the wealth repercussions originating from the opening or the closure of whole market sectors: in this context, for example, the application of the measures could result in the generation of innovative interpretations and solutions of present major geopolitical questions, including the international turmoil consequent to the invasion of Ukraine by Russia.

8 Appendix A: Convergent cross-mapping method

Identifying causality in complex systems can supply a deeper understanding of their internal structure and allows to generate more reliable predictions on their future global evolution. On the practical level, these implications play in turn a fundamental role in the regulation of many professional and scientific environments, such as ecology, epidemiology, finance and many others. However, the search for causality is considerably complicated by the ubiquity of nonlinear and chaotic patterns in such systems, which often hinder the detection of unambiguous causality signals.

Although it provides neither a sufficient nor a necessary condition for causality, the dynamical correlation between the variables of a complex system represents the most employed heuristic measure of the casual relation potentially existing between them. However, infering causality from simple correlation can lead to incorrect conclusions in many scenarios. For example, even in the simplest nonlinear systems, it is not unusual for a couple of variables that have presented a positive dynamical correlation for a long period to suddenly appear decoupled or even evolve in anti-correlated trends, generating issues in the fitting of the data and contradictions in their interpretation.

From a theoretical standpoint, the paradigm proposed by Sugihara in [30], named *convergent cross-mapping method* (CCM) for technical reasons, makes use of a particular result of embedding theory in order to employ the dynamical correlation between two variables as a measure of the likelihood of their causal connection. More specifically, the method provides a couple of necessary conditions for the validity of causal relations of different kind and allows to discard the scenarios which cannot contemplate their existence. From the application point of view, CCM is specifically designed for the description of chaotic systems characterized by weak to moderate couplings between their components, which usually cannot be addressed through the application of other paradigms (like the celebrated Granger causality).

The next two sections present, respectively, the theoretical background which subtends the application of the method and a thorough description of the algorithm which performs the evaluation of the dynamical correlation.

8.1 Theoretical background

The application of delay embedding theorems has encountered increasingly widespread approval during the last decades, due to their usefulness in experimental contexts. In general, a delay embedding theorem provides a peculiar theoretical framework for the description of the dynamics of a physical system and relates it to delayed time series of one or more proper functions. The first result in this field was provided in 1981 by Takens, who formulated and applied the first delay theorem inside a fluid dynamics study [43]. As shown by the following statement, the theorem describes the evolution of the system along the lines of classic dynamics theory:

THEOREM 1 (Takens's theorem): Let M be a compact attractor manifold of dimension d. Let X be a smooth vector field on M with flow ϕ: $(M, \mathbb{R}^+) \to M$ and let f: $M \to \mathbb{R}$ be a smooth function on M, where "smooth" means at least of

class C^2. Let $\tau > 0$ and let ϕ_τ: $M \to M$ be the fixed-time flow of X. Then, the map $\Phi_{X,f}$: $M \to \mathbb{R}^{2d+1}$ defined by $\Phi_{X,f}(x) = \big(f(x), f(\phi_\tau(x)), ..., f(\phi_\tau^{2d}(x))\big)$, where $\phi_\tau^n \equiv \underbrace{\phi_\tau \circ \cdots \circ \phi_\tau}_{n \text{ times}}$, is an embedding of M in $\mathbb{R}^{2d+1}$.

The appeal of this basic result mainly lies in the strong generality of its assumptions; however, a relevant limitation to its scope comes from the pronounced separation between the dimension of the attractor manifold and the one of the embedding space. In subsequent years, many extensions of the theorem have been proposed in a number of independent studies, with the purpose of broadening its scope by either generalizing the attractor manifold [44] or extending the space of feasible functions [45]. The particular delay embedding theorem which subtends the development of CCM is part of a study [46] which presents several results along both the branches of research. The statement that follows represents a slightly weaker version of the theorem, better suited to the purposes of this discussion:

THEOREM 2 (Fractal Takens's theorem): Let M be a compact strange attractor manifold of dimension d and fractal dimension d'. Let X be a smooth vector field on M with flow ϕ: $(M, \mathbb{R}^+) \to M$ and let f: $M \to \mathbb{R}$ be a smooth function on M, where "smooth" means at least of class C^1. Let $\tau > 0$ and let ϕ_τ: $M \to M$ be the fixed-time flow of X. Let $2d' < m \leq 2d + 1$ be an integer. Assume that M contains at most a finite number of equilibria of ϕ, no periodic orbits of period τ or 2τ and at most a finite number of periodic orbits of period $n\tau$ ($n > 2$) and that the linearization of any periodic orbit has distinct eigenvalues. Then, the map $\Phi_{X,f}$: $M \to \mathbb{R}^m$ defined by $\Phi_{X,f}(x) = \big(f(x), f(\phi_\tau(x)), ..., f(\phi_\tau^{m-1}(x))\big)$, where $\phi_\tau^n \equiv \underbrace{\phi_\tau \circ \cdots \circ \phi_\tau}_{n \text{ times}}$, is an embedding of M in $\mathbb{R}^m$.

Actually, as proved in [46], the same conclusion also holds if the attractor manifold is replaced by a generic compact real subset; moreover, the theorem can be strengthened by implementing the concept of *prevalence*, an infinite-dimensional extension of the *almost everywhere* concept of measure theory which allows to deepen the characterization of the space of the functions which satisfy the theorem. Still, the above result implies that by slightly enriching the description of the attractor manifold through the embodiment of a fractal structure and by choosing a low fractal dimension (which is not necessarily an integer), the aforementioned separation from the actual dimension of the manifold and the dimension of the embedding space can be strongly reduced: in particular, the most relevant consequence of this fact is the possibility to implement minimal embedding spaces without strongly limiting the freedom of motion of the system. Remarkably, in many contexts, the addition of the fractal hypothesis does not represent a relevant limitation to the scope of the theorem: very often, the dynamics of chaotic systems (like a turbulent fluid or a country's economy) can indeed be optimally described through the implementation of attractor manifolds with a strange nature. Furthermore, the additional requests concerning the absence of periodic orbits can be

easily satisfied by choosing a sufficiently small value for the time lag τ.

Considering the bijective relation between the attractor manifold and the embedded ones, a direct consequence of Theorem 2 is the possibility to perform a composition between the embeddings related to two distinct smooth functions, in order to create a one-to-one correspondence between their time series. In particular, the practical usefulness of this result lies in the opportunity to gain an insight on the dynamical nature of the observables of the system, represented on the technical level by the aforementioned functions. The requirement on the smoothness of an observable function is equivalent to characterize it as a variable dependent on the dynamics of the system; conversely, the opposite condition is peculiarly satisfied by variables which are not dynamically influenced by the system itself (this is the case, for example, of an external forcing one). Moreover, if two distinct functions satisfy the smoothness condition, their patterns along the orbits of the attractor manifold are dynamically coupled, because both the variables represented exercise an effective influence on the evolution of the other: in this case, the two functions are said to *cross-map* each other.

8.2 Symplectic algorithm

CCM specifically provides information about the nature of the coupling that potentially binds the dynamics of the variables. Practically, the method involves the implementation of an algorithm for the evaluation, through cross-mapping operations, of the correspondence between the respective empirical time series, in order to assess the likelihood of the diffeomorphisms described in Theorem 2. It should be noted that, as already clear from the discussion in Section 6.2, the precise definition of the system and of the attractor manifold which contains its motion (in particular, of its fractal dimension) plays a marginal role in the context of the practical analysis of the empirical time series.

Delving into the technical details, the algorithm is based on simplex projection ([47], [48]), a method which performs cross-map reconstructions of the single records by considering exponentially weighted distances between nearby time-lagged vectors. Adopting the same notation seen in section 6.2, the variables involved are denoted with X and Y, while M_X and M_Y denote the respective embedded manifolds. In the case of X, the empirical time series obtained from a single experiment, indicated with $\{X(t_n)\}_{n=0,...,L}$ (for $L > 0$), offers a reconstruction of a certain trajectory on M_X; in particular, the generic time-lagged vector crossed by the estimated trajectory has the form $(X(t_n), X(t_n + \tau), ..., X(t_n + (E-1)\tau))$ and is indicated with $\vec{T}_X(t_n)$, where the (positive) time lag τ is a multiple of the interval between the records and E is the dimension of the selected embedding space. The first step of the cross-map reconstruction of $\vec{T}_X(t_n)$ consists in locating the simultaneous vector on M_Y (namely, $\vec{T}_Y(t_n)$) and finding its $E+1$ nearest neighbors; this particular value is equal to the minimum of points needed for a bounding simplex in an E-dimensional space [47]. For the sake of clarity, it is convenient to relabel the time indices of the nearest neighbors as $\{t'_m\}_{m=0,...,E}$, where increasing values of m represent increasing distances from $\vec{T}_Y(t_n)$: for example, the nearest and the furthest neighbors are now respectively indicated with $\vec{T}_Y(t'_0)$ and $\vec{T}_Y(t'_E)$. In turn, the Euclidean distance of the m-th closest neighbor is denoted with d_m. Through simplex projection, the estimate of $\vec{T}_X(t_n)$ can

be simply translated into an estimate of its first component, which is provided by the following weighted average:

$$X(t_n) \approx \sum_{m=0}^{E} w_m X(t'_m) \equiv \tilde{X}(t_n) \qquad , \tag{8.1}$$

where the weights embody the aforementioned distances and are given by

$$w_m = \frac{u_m}{\sum_{l=0}^{E} u_l} \quad \text{with} \quad u_m = \mathrm{e}^{-\frac{d_m}{d_0}} \quad . \tag{8.2}$$

As shown, the nearest neighbors on M_Y are mapped to their simultaneous counterparts on M_X, whose first components practically supply the desired estimation: in particular, the closer a vector is to $\vec{T}_Y(t_n)$, the greater the relevance of its counterpart on M_X in the estimation of $\vec{T}_X(t_n)$. The error on $\tilde{X}(t_n)$ simply consists in the standard deviation of the weighted average in Eq. 8.1 and reflects the precision of the localization of $\vec{T}_X(t_n)$ on M_X. Obviously, the same considerations hold also in the case of the reconstruction of the records of Y starting from M_X. It should be noted that, given the structure of the time-lagged vectors, the procedure described can be employed for the reconstruction of a minimum of $E+2$ records (due to the minimum number of nearest neighbors) and a maximum of $L-E+1$ records of a time series.

At this point, the matching degrees between the empirical and the reconstructed time series can be assessed in several ways: the simplest method, employed in all the case studies presented in [30] and in Section 6.2, consists in the evaluation of the Pearson's correlation coefficient, whose global error is simply provided by the propagation of the ones of the single records. In any case, in light of the theoretical results presented previously, the basic application of CCM to the time series in their totality not only provides a measure of the correlation between the dynamics of the variables of the system (through the multiple identifications $X(t_n) \leftrightarrow \vec{T}_Y(t_n)$ and $Y(t_n) \leftrightarrow \vec{T}_X(t_n)$, $\forall n$), but also allows to verify a necessary condition for the validity of Theorem 2: if the assumptions of the latter are satisfied for variable X, a pronounced matching degree between the empirical and the reconstructed time series related to Y must indeed be detected and vice versa. In addition, the power of CCM can be further enhanced by studying the evolution of the matching degrees as larger fractions of the time series are implemented in the analysis, starting from the minimum number of records mentioned before. In this scenario, the monotonic convergence of a matching degree towards an asymptotic value represents another necessary consequence of Theorem 2: indeed, the increase of the number of time-lagged vectors on M_X and the consequent filling of the related trajectory must induce, if the assumptions of the theorem are satisfied, a progressive improvement of the reconstruction of the records of Y, and vice versa. In the end, the peculiar usefulness of CCM precisely lies in the possibility to test both these necessary conditions, in order to gain an insight on the dynamical nature of the variables considered and on the coupling potentially binding them.

Lastly, it should be emphasized that, in an experimental context, the final results provided by CCM are sensibly affected by a series of limiting factors (including obser-

vational errors, process noise and limited length of the time series) and then admit, at most, a probabilistic interpretation. For example, a strongly and monotonically convergent matching degree between the time series related to X signals the likely presence of the dynamical correlation which is necessary in order to describe the dynamics of Y as a function of the states of X itself; in addition, the detection of similar matching degrees for both the variables opens the possibility to describe their dynamics through a bidirectional coupling. Conversely, if the matching degree related to X evidently fails to satisfy even only one of the two necessary conditions, very probably Y cannot be described in terms of a smooth function on the attractor manifold and its dynamics is then independent of the states of X itself (for example, the former could be interpreted as a forcing external variable). Obviously, in any intermediate scenario the influence of the aforementioned limiting factors prevents an univocal interpretation of the final results.

References

[1] Ma Angeles Serrano and Marián Boguná. Topology of the world trade web. *Physical Review E*, 68(1):015101, 2003.

[2] R Haussman and B Klinger. Structural transformation and patterns of comparative advantage in the product space. *Center for International Development Working Paper*, (128), 2006.

[3] Andrea Tacchella, Matthieu Cristelli, Guido Caldarelli, Andrea Gabrielli, and Luciano Pietronero. A new metrics for countries' fitness and products' complexity. *Scientific reports*, 2(1):1–7, 2012.

[4] Michele Caraglio, Fulvio Baldovin, and Attilio L Stella. Export dynamics as an optimal growth problem in the network of global economy. *Scientific reports*, 6(1):1–10, 2016.

[5] Jan Tinbergen. Shaping the world economy; suggestions for an international economic policy. 1962.

[6] David R Nelson and Nadav M Shnerb. Non-hermitian localization and population biology. *Physical Review E*, 58(2):1383, 1998.

[7] Jean-Philippe Bouchaud and Mark Potters. Theory of financial risks and derivative pricing, 2004.

[8] A-L Barabási, Harry Eugene Stanley, et al. *Fractal concepts in surface growth*. Cambridge university press, 1995.

[9] Timothy Halpin-Healy and Yi-Cheng Zhang. Kinetic roughening phenomena, stochastic growth, directed polymers and all that. aspects of multidisciplinary statistical mechanics. *Physics reports*, 254(4-6):215–414, 1995.

[10] Jean-Philippe Bouchaud and Marc Mézard. Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and its Applications*, 282(3-4):536–545, 2000.

[11] Thomas Gueudré, Alexander Dobrinevski, and Jean-Philippe Bouchaud. Explore or exploit? a generic model and an exactly solvable case. *Physical review letters*, 112(5):050602, 2014.

[12] Jonathan D Cohen, Samuel M McClure, and Angela J Yu. Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):933–942, 2007.

[13] James G March. Exploration and exploitation in organizational learning. *Organization science*, 2(1):71–87, 1991.

[14] Gianluca Teza, Michele Caraglio, and Attilio L Stella. Growth dynamics and complexity of national economies in the global trade network. *Scientific reports*, 8(1):1–8, 2018.

[15] Gianluca Teza, Michele Caraglio, and Attilio L Stella. Entropic measure unveils country competitiveness and product specialization in the world trade web. *Scientific reports*, 11(1):1–11, 2021.

[16] César A Hidalgo and Ricardo Hausmann. The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26):10570–10575, 2009.

[17] Matthieu Cristelli, Andrea Gabrielli, Andrea Tacchella, Guido Caldarelli, and Luciano Pietronero. Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one*, 8(8):e70726, 2013.

[18] Matthieu Cristelli, Andrea Tacchella, and Luciano Pietronero. The heterogeneous dynamics of economic complexity. *PloS one*, 10(2):e0117174, 2015.

[19] Emanuele Pugliese, Andrea Zaccaria, and Luciano Pietronero. On the convergence of the fitness-complexity algorithm. *The European Physical Journal Special Topics*, 225(10):1893–1911, 2016.

[20] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[21] Ian F Spellerberg and Peter J Fedor. A tribute to claude shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'shannon–wiener'index. *Global ecology and biogeography*, 12(3):177–179, 2003.

[22] Pier Paolo Saviotti and Koen Frenken. Export variety and the economic performance of countries. *Journal of Evolutionary Economics*, 18(2):201–218, 2008.

[23] Oktay K Pashaev. Quantum coin flipping, qubit measurement, and generalized fibonacci numbers. *Theoretical and Mathematical Physics*, 208(2):1075–1092, 2021.

[24] Angus Maddison. *A millennial perspective*. OECD, 2001.

[25] Lant Pritchett. Divergence, big time. *Journal of Economic perspectives*, 11(3):3–17, 1997.

[26] Adam Smith. An inquiry into the nature and causes of the wealth of nations: Volume one. London: printed for W. Strahan; and T. Cadell, 1776., 1776.

[27] Paul M Romer. Endogenous technological change. *Journal of political Economy*, 98(5, Part 2):S71–S102, 1990.

[28] Gene M Grossman and Elhanan Helpman. Quality ladders in the theory of growth. *The review of economic studies*, 58(1):43–61, 1991.

[29] Gianluca Teza, Michele Caraglio, and Attilio L Stella. Data driven approach to the dynamics of import and export of g7 countries. *Entropy*, 20(10):735, 2018.

[30] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *science*, 338(6106):496–500, 2012.

[31] http://cepii.fr/CEPII/en.

[32] http://cepii.fr/CEPII/en/publications/publicat.asp.

[33] http://comtrade.un.org.

[34] http://wcoomd.org.

[35] http://data.worldbank.org.

[36] David S Watkins. *Fundamentals of matrix computations*. John Wiley & Sons, 2004.

[37] D Glasser, FJM Horn, and R Meidan. Properties of certain zero column-sum matrices with applications to the optimization of chemical reactors. *Journal of Mathematical Analysis and Applications*, 73(2):315–337, 1980.

[38] AJ Roberts. Modify the improved euler scheme to integrate stochastic differential equations. *arXiv preprint arXiv:1210.0933*, 2012.

[39] R Bruce Kellogg, Tien-Yien Li, and James Yorke. A constructive proof of the brouwer fixed-point theorem and computational results. *SIAM Journal on Numerical Analysis*, 13(4):473–483, 1976.

[40] César A Hidalgo. The dynamics of economic complexity and the product space over a 42 year period. *CID Working Paper Series*, 2009.

[41] Kadri Kuusk and Mikhail Martynovich. Dynamic nature of relatedness, or what kind of related variety for long-term regional growth. *Tijdschrift voor economische en sociale geografie*, 112(1):81–96, 2021.

[42] Sándor Juhász, Tom Broekel, and Ron Boschma. Explaining the dynamics of relatedness: The role of co-location and complexity. *Papers in Regional Science*, 100(1):3–21, 2021.

[43] F Takens. Lecture notes in mathematics, dynamical system and turbulence. *DA Rand & LS Young, Springer, New York*, 1981.

[44] James C Robinson. A topological delay embedding theorem for infinite-dimensional dynamical systems. *Nonlinearity*, 18(5):2135, 2005.

[45] Ethan R Deyle and George Sugihara. Generalized theorems for nonlinear state space reconstruction. *Plos one*, 6(3):e18295, 2011.

[46] Tim Sauer, James A Yorke, and Martin Casdagli. Embedology. *Journal of statistical Physics*, 65(3):579–616, 1991.

[47] George Sugihara and Robert M May. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344(6268):734–741, 1990.

[48] Owen L Petchey. Simplex projection walkthrough. 2016.