Physica A 387 (2008) 6595–6604

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Globalization — Entropy unification through the Theil index

Janusz Miśkiewicz *

Institute of Theoretical Physics, University of Wrocław, pl. M. Borna 9, 50-204 Wrocław, Poland

ARTICLE INFO

Article history:
Received 16 June 2008
Received in revised form 11 July 2008
Available online 8 August 2008

PACS:
05.45.Tp
89.65.Gh

Keywords:
Econophysics
Time series analysis
Globalization
Entropy
Network
Minimum spanning tree

ABSTRACT

A time series is remapped onto an entropy concept, based on the Theil index. The Manhattan distance between these surrogate series is calculated, and contrasted to the usual correlation distance measure. The idea is applied to several Gross Domestic Product (relative increments) of rich countries. Such distances are calculated for various time window sizes. The role of time averaging in such finite size windows is discussed. We construct the locally minimum spanning tree (LMST) corresponding to the distance matrix. Another hierarchical network structure (Unidirectional Minimal Length Path) is compared with the LMST for confirming that the mean distance between the most developed countries on different networks actually decreases in time, — which we consider as a proof of economy globalization. It is stressed that this *entropy distance measure* seems more suitable in detecting some "phase transition" in time series, like a globalization process than the usual correlation based measure.

1. Introduction

Let us define the vocabulary and the topics of interest before going into a detailed discussion:

Definition 1. Globalization — the process by which diffusion of *commodities* and *ideas*, can foster a standardization of "cultural" expressions around the world [1].

We restrict this definition here below to the case of *economy* in order to get a quantitative aspect.

Definition 2. Globalization process in *economy* is understood as the increase of similarities in development patterns.

How to measure *similarities* and which (economy) parameters have to be taken into consideration in searching for globalization features are rather important questions which arise from various points of view: going from mathematics to politics. Much debate exists, — many admit that the different approaches might intrinsically imply a certain type of conclusion. The similarities are here below considered to be measured through *distances* between time series of classical economic parameters. Within this paper, the Gross Domestic Product (GDP) has been selected as the most often representative parameter describing the status of an economy, since *mutatis mutandis* it is defined for all countries, — Section 2. We stress at once that we prefer to discuss GDP rather than GDP/capita, the latter being dependent on factors other than strict economy measures, e.g. migration policy, health care, life span, etc. However the GDP *relative increments* are still better parameters for this sort of investigation; see related and interesting discussions in economy publications, e.g. Refs. [2–5].

* Tel.: +48 603918659.
E-mail address: jamis@ift.uni.wroc.pl.

0378-4371/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.08.007

For simplicity here below we consider the distance between two time series $A(t)$ and $B(t)$ for country A and B as defined through the Minkowski distance of order p (p-norm distance), particularly the 1-norm distance

$$d_{(A,B)}(t,T) = \sum_{i=t-T}^{t} |A(t) - B(t)|. \tag{1}$$

However since one aims at comparing the disorder levels in the GDP evolution, we first map each GDP series onto another, taking an *entropy* point of view. The latter is based on the so called Theil index (Eq. (2)) [6]

$$E_A(t,T) = \sum_{i=t-T}^{t} \left(\frac{A(i)}{\sum_{j=t-T}^{t} A(j)} \ln \frac{A(i)}{\langle A \rangle_{(t,T)}} \right) \tag{2}$$

where $\langle A \rangle_{(t,T)} = \frac{1}{T} \sum_{j=t-T}^{t} A(j)$ in the above Eq. (1).

Within this framework one can define the so called *entropy distance* between two time series as,

$$d^{(E)}_{(A,B)}(t,T) = |E_A(t,T) - E_B(t,T)| \tag{3}$$

which rather defines the local level of disorder, rather than the global one as in the usual Shannon entropy definition, $S \simeq \sum_i p_i \ln p_i$, where i labels some system state occurring with probability p_i. In practice [6] the Theil index measures the difference between the system maximum entropy and its entropy at a given time. The Theil index is usually used to compare economies, its features or grouping structures [7–9]. It has not been used, it seems, to define a distance between the presently investigated sets.

For smoothing the role of likely "political and other noise" in such series, a time averaging is made over different and possible size windows; the range of optimal ones is discussed from this entropy distance measure point of view through the data of interest. After constructing the various matrix distances between "countries" (a short hand notation for "GDP relative increments between countries"), we propose to examine the statistical properties of such distances through network structures. The network methods are usually very meaningful in analyzing relationship between subjects e.g. Refs. [10, 11] and others. The definition of applied networks i.e. the locally minimum spanning tree (LMST) and the unidirectional minimum length path (UMLP) are given in Section 2.2. The results are found in Section 3.

We consider 21 time series, i.e. most of the richest countries. The data spans about 60 years, 1950–2005. Moreover, it is shown that the mean distance between the most developed countries on both networks decreases in time, surely since the 90's, — which we consider is a proof of globalization. It is stressed that the entropy distance measure seems more suitable in detecting a phase transition in noisy time series of complex non equilibrium systems, like a globalization process, than the usual correlation based measure [12].

2. Description of the system

2.1. GDP data

The total GDP data collected by the Gröningen Growth and Development Center [13] is used as the input data. The database provides data normalised along the Elteto–Köves–Szulc (EKS) or Geary-Khamis (GK) aggregation process. A description and analysis of the properties of these aggregation methods can be found in e.g. Refs. [14–16]. Much debate exists in economy circles about the quality of such GDP data *per se* and for defining an economy. Such a discussion is out of place here.

The series here below studied are those corresponding to the 21 most developed nations[1]: Austria (AT), Belgium (BE), Denmark (DK), Finland (FI), France (FR), Germany (DE), Greece (GR), Iceland (IS), Ireland (IR), Italy (IT), Luxembourg (LU), The Netherlands (NL), Norway (NO), Portugal (PT), Spain (ES), Sweden (SE), Switzerland (CH), United Kingdom (UK), Canada (CA), USA (US), Japan (JP). We also include an artificial country "All" which is taken as a benchmark to describe an average level of development. Specifically here the GDP of the "All" is constructed as the unweighted sum of the 21 considered country GDP's. The recorded time interval spans 55 years, i.e. from 1950 till 2005.

The statistical properties of the data have been analysed elsewhere [12]. Let us recall the most important findings. In the analysis, the mutual relationship between countries is considered through a focus on country hierarchical rankings. The following rankings were discussed in Ref. [12]: time integral of GDP, the mean, standard deviation and mean/standard deviation of the GDP relative increments. The results of the subsequent ranking and basic statistical properties are similar for both EKS and GK aggregation methods. Also the basic statistical properties of the GDP ratio, say at time t, for country A, defined by

$$\Delta GDP_A(t) = \frac{GDP_A(t+1)}{GDP_A(t)} \tag{4}$$

[1] The country name abbreviations are following the internet top-level domains naming scheme.