

A moving-average-minimal-length-path method for finding EU country clusters

by M. Gligor^{1,2} and M. Ausloos²

¹National Coll. 'Roman Voda', Romania

²SUPRATECS, ULg, Euroland

“Science is built of facts the way a house is built of bricks; but an accumulation of facts is no more science than a pile of bricks is a house”

Henri Poincaré, La science et la hypothèse

CONTENT

- Introduction
- MAMLP method
- The 'globalization' seen as an after-shock relaxation process
- The evidences of globalization in 5 years time window size
- Clustering of EU-15 countries by MAMLP method
- Further developements

The usual ways to express the **correlation** between two **time series**:

- The correlation coefficient:

$$C(x, y) = \frac{\langle x_i y_i \rangle - \langle x_i \rangle \langle y_i \rangle}{\sqrt{\langle x_i^2 - \langle x_i \rangle^2 \rangle \langle y_i^2 - \langle y_i \rangle^2 \rangle}}$$

- The statistical distance:

$$d(x, y) = \sqrt{2 \cdot [1 - C(x, y)]}$$

Minimum Spanning Tree

(O. Borukva, 1926; J. Kruskal, 1956; R.C. Prim, 1957)

- Borukva's algorithm - an efficient electrical coverage of Bohemia
- Kruskal's algorithm - find a subset of the edges that forms a tree that includes every vertex, where the total weight of all the edges in the tree is minimized
- Prim's algorithm – MST for a connected weighted graph
- All three are **greedy algorithms** – making the locally optimum choice at each stage with the hope of finding the global optimum – running in **polynomial time**

Minimum Spanning Tree

(R. Mantegna, 1999; Bonnano et al., 2000)

- The time series → The stock price variations
- The statistical distances become the elements of the distance matrix $D \Rightarrow$ a topological arrangement between the different stocks
- The use of the **ultrametric subdominant space** and of the ultrametric distance between objects \Rightarrow clusters of the stocks according to the industry they belong to.

The macroeconomic (ME) indicators analyzed and the data sources

- **The Gross Domestic Product / capita**
- **The Gross Domestic Product (GDP)**
- **The Final Consumption Expenditure**
- **The Gross Capital Growth**
- **The Net Exports (Exports – Imports)**

<http://devdata.worldbank.org/query/default.htm>

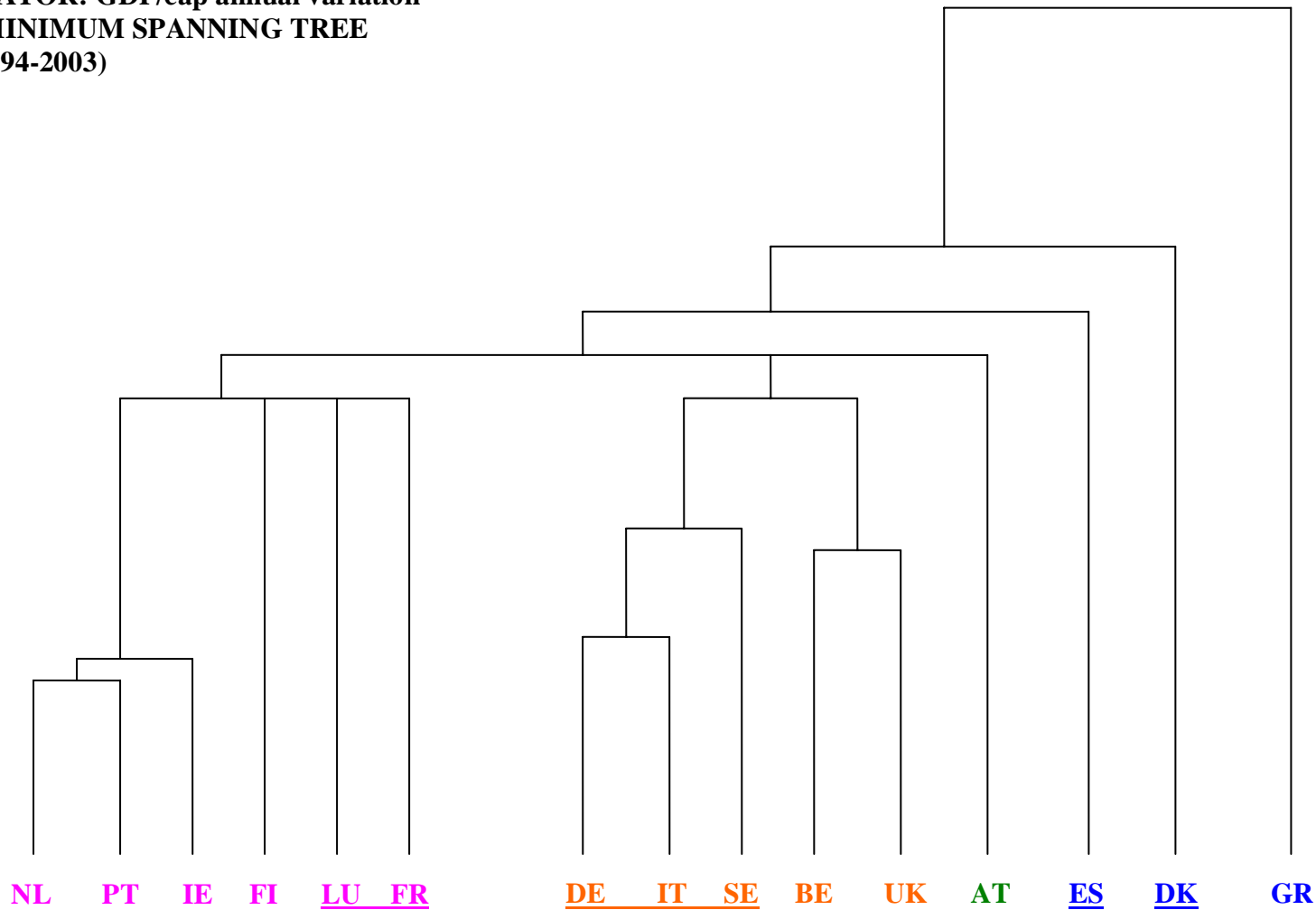
<http://www.cia.gov/cia/publications/factbook/rankorder/2004rank.html>

http://www.oecd.org/countrieslist/0,3025,en_33873108_33844430_1_1_1_1_1,00.html

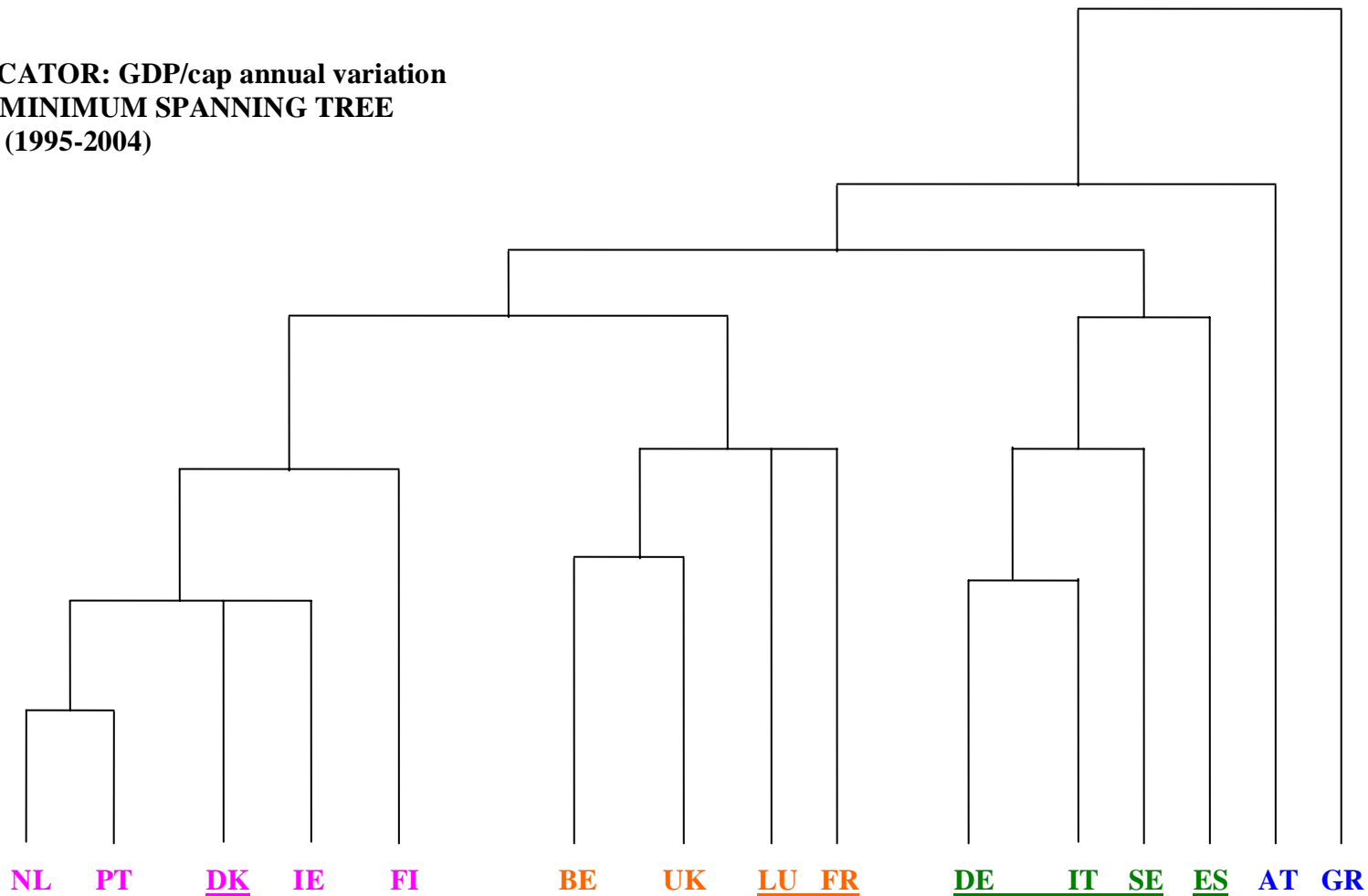
The Minimal Spanning Tree and the macroeconomic time series

- Unlike the high frequency financial data series, the ME time series are **sparse, short and noisy**.
- Most of data is binned into ranges of values, so they are subject to **gross approximations** and dependent on **sampling methodology**.
- When too small differences between the linkage distances arise, MST **is not unique**.
- The tree-like structure **is not robust** against fluctuations induced by a moving time window.

INDICATOR: GDP/cap annual variation
THE MINIMUM SPANNING TREE
(1994-2003)



INDICATOR: GDP/cap annual variation
THE MINIMUM SPANNING TREE
(1995-2004)



LMST and the MLP Algorithms

(Miskiewicz and Ausloos, 2005a,b)

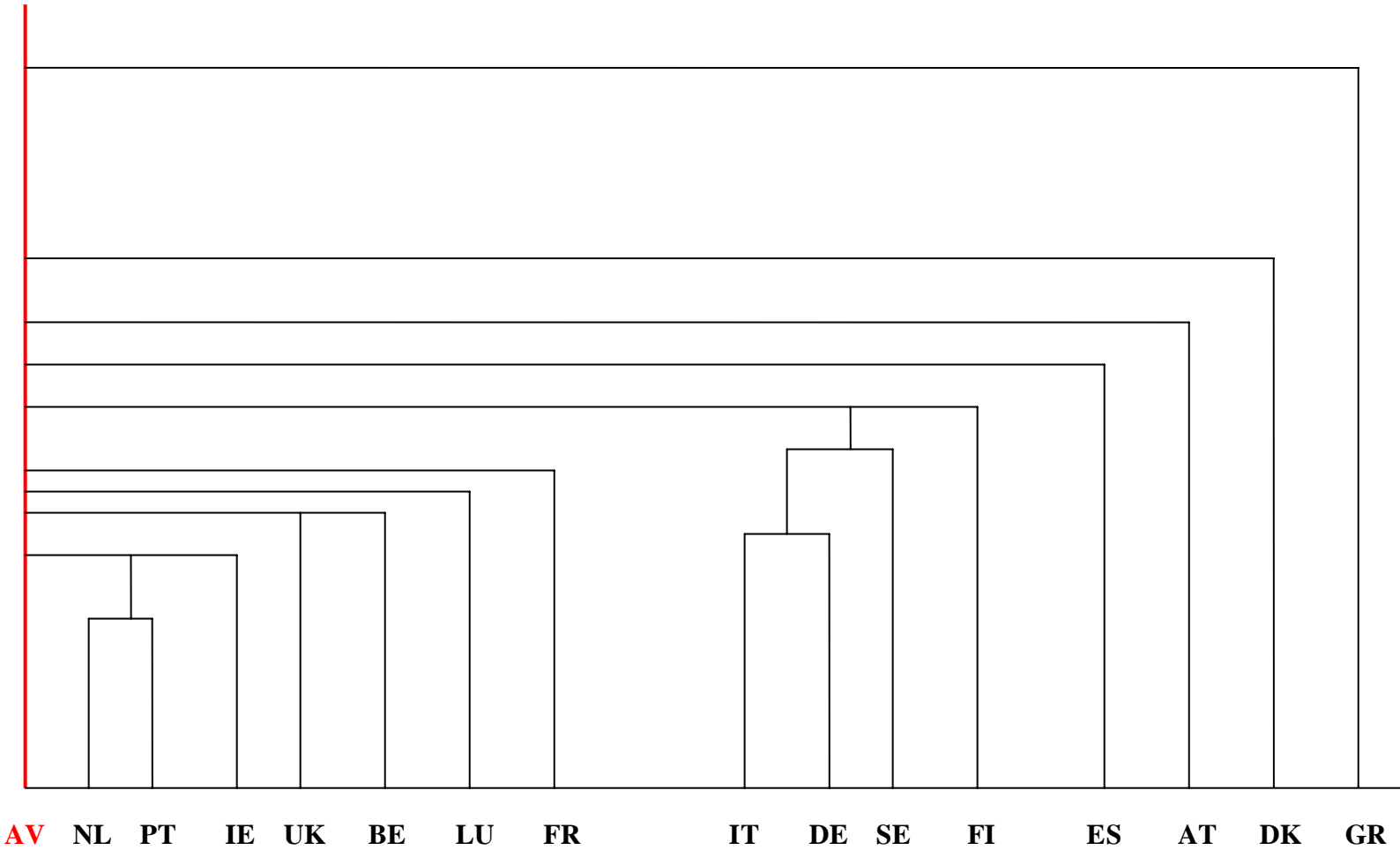
- **Local Minimum Spanning Tree:** the initial pair of nodes (the root) of the tree is the pair with the strongest correlation
- **Minimal Length Path Algorithms:** the closest neighbouring countries are attached at the end of a chain
- Some **arbitrariness** in the root of the tree: the “**All**” country as a more **common root** from which to let the tree grow was permitting a better comparison.

The Moving Average Minimal Length Path (MAMLP) Method

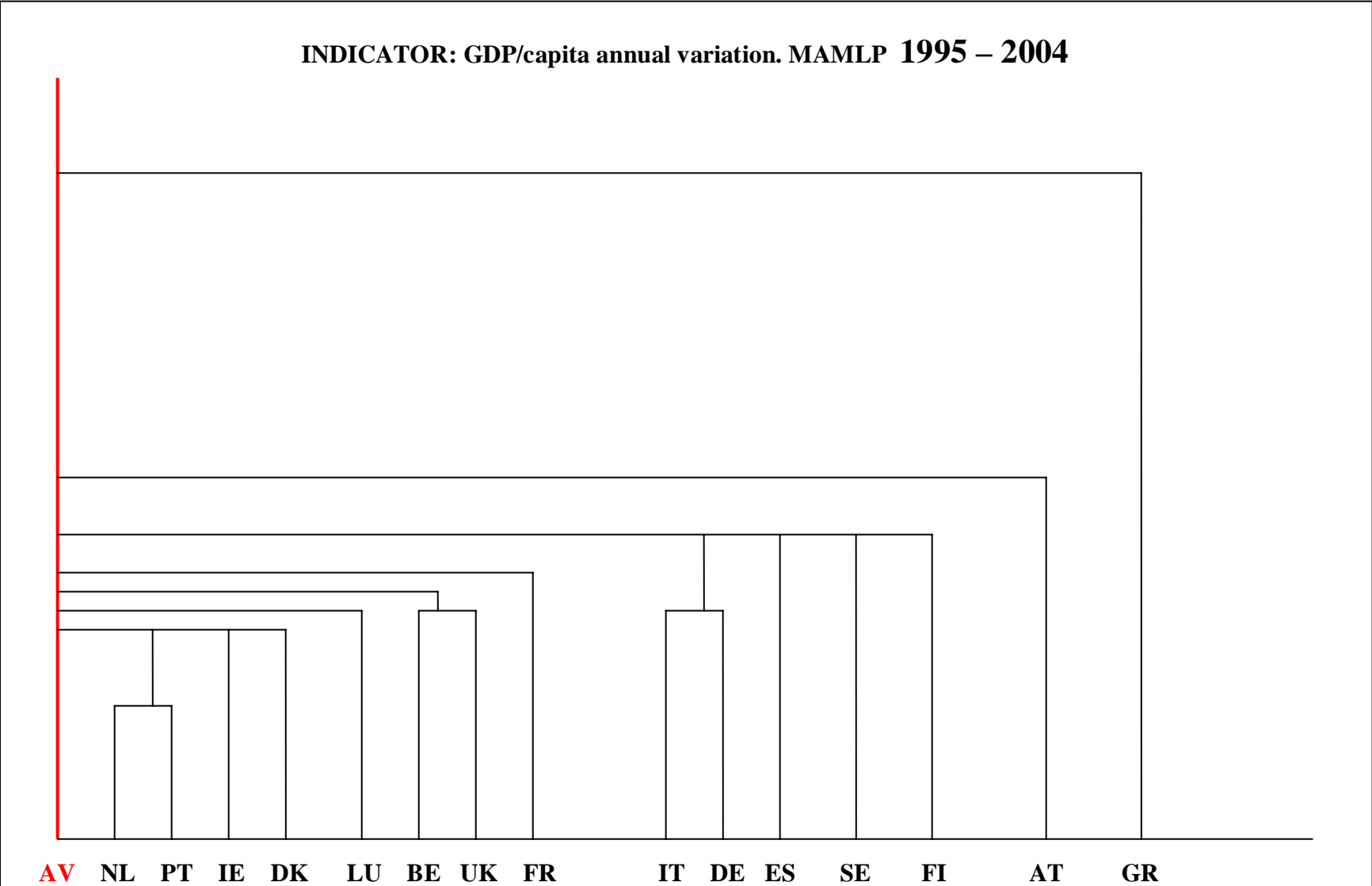
The moving-average minimal-length-path (MAMLP) method

- 1) An 'average' agent (**AV**) is virtually included into system;
- 2) The statistical distance matrix is constructed, and thereafter, the elements are set into increasing order
- 3) The hierarchy is constructed, connecting each agent by its **minimal path length to AV**. Its minimal distance to AV is associated to each agent.

INDICATOR: GDP/capita annual variation. MAMLP 1994 – 2003



INDICATOR: GDP/capita annual variation. MAML 1995 – 2004



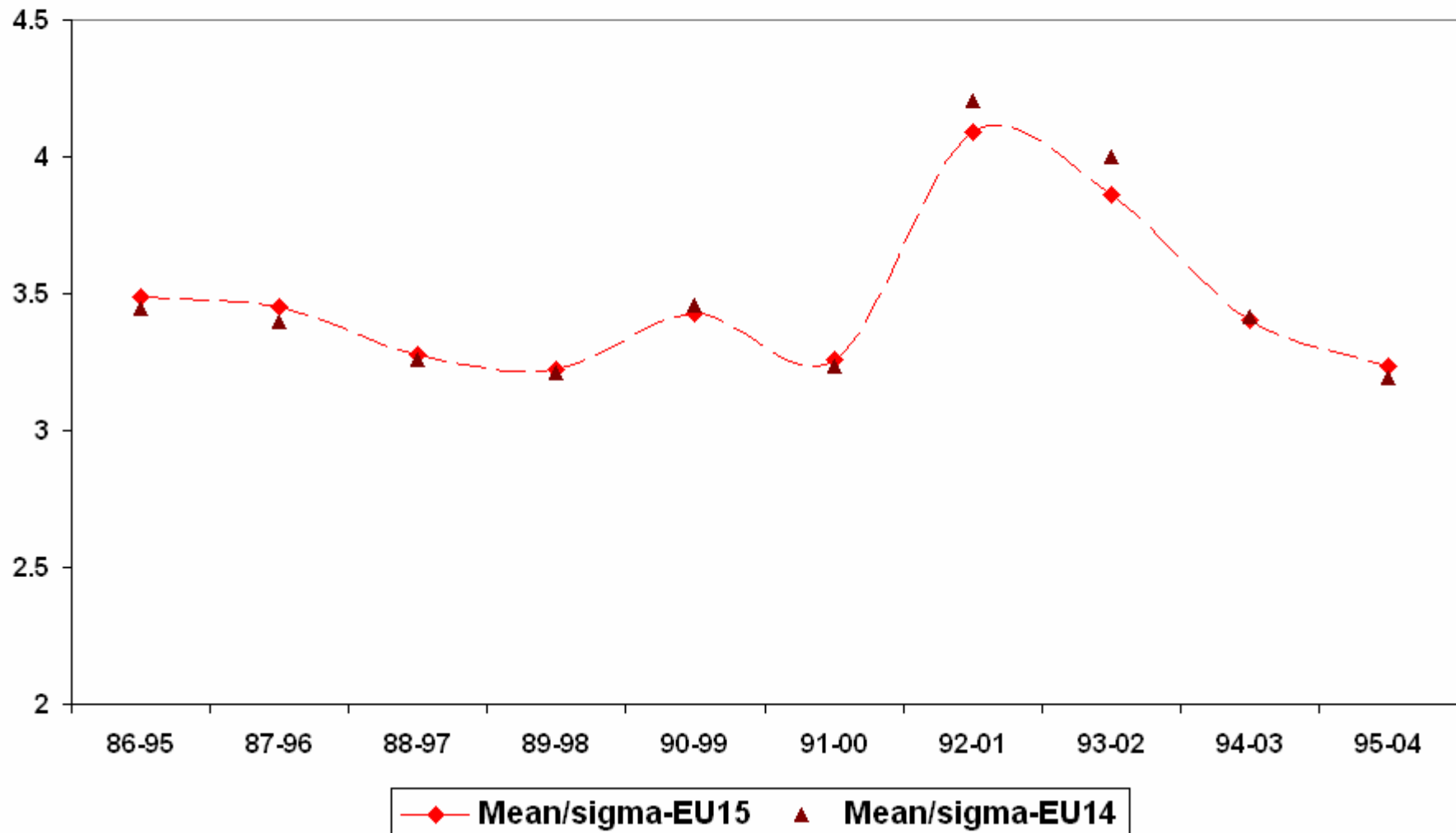
The moving-average minimal-length-path (MAMLP) method

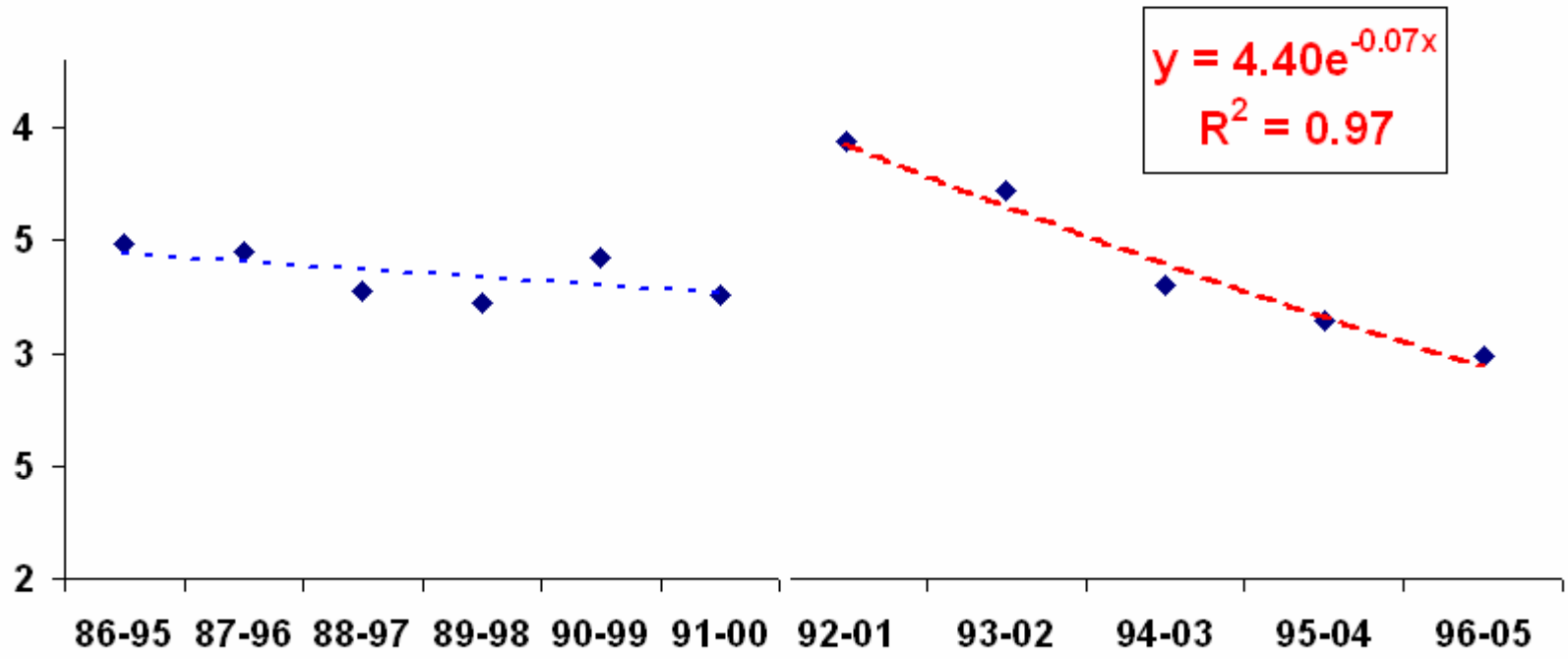
- 4) The procedure is repeated by **moving** a given and constant time window over the investigated time span.
- 5) The statistical properties of the datasets are investigated.
- 6) The agents are sorted through their movement inside the hierarchy. A new **correlation matrix** between country distances to their mean is constructed.

RESULTS (1)

The “globalization” seen as an
“aftershock” relaxation process

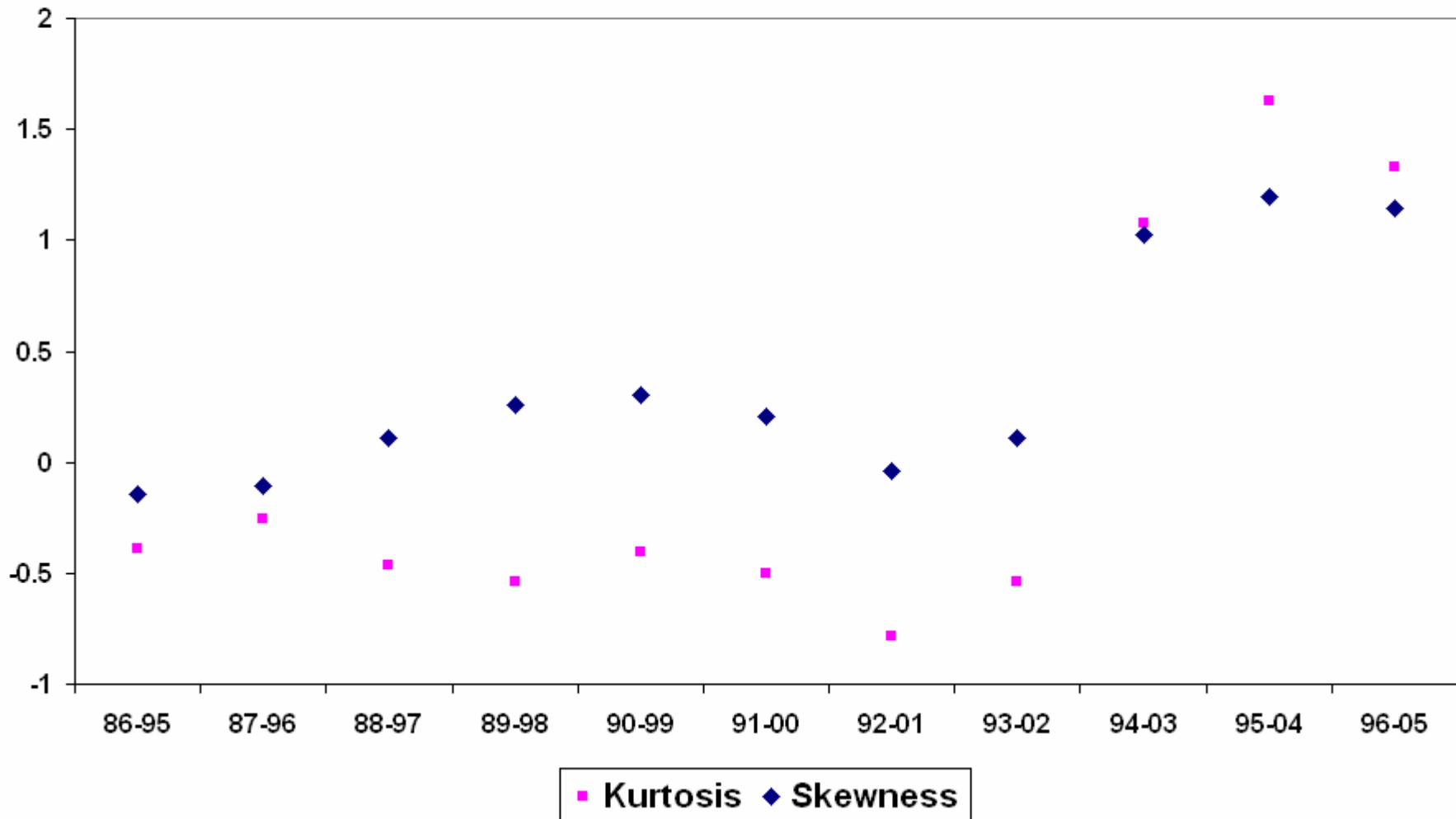
**The time evolution of the mean statistical distance between EU-15.
Indicator: GDP/capita. Time Window Size: 10 years**





Statistical properties of the EU-15 distances distribution

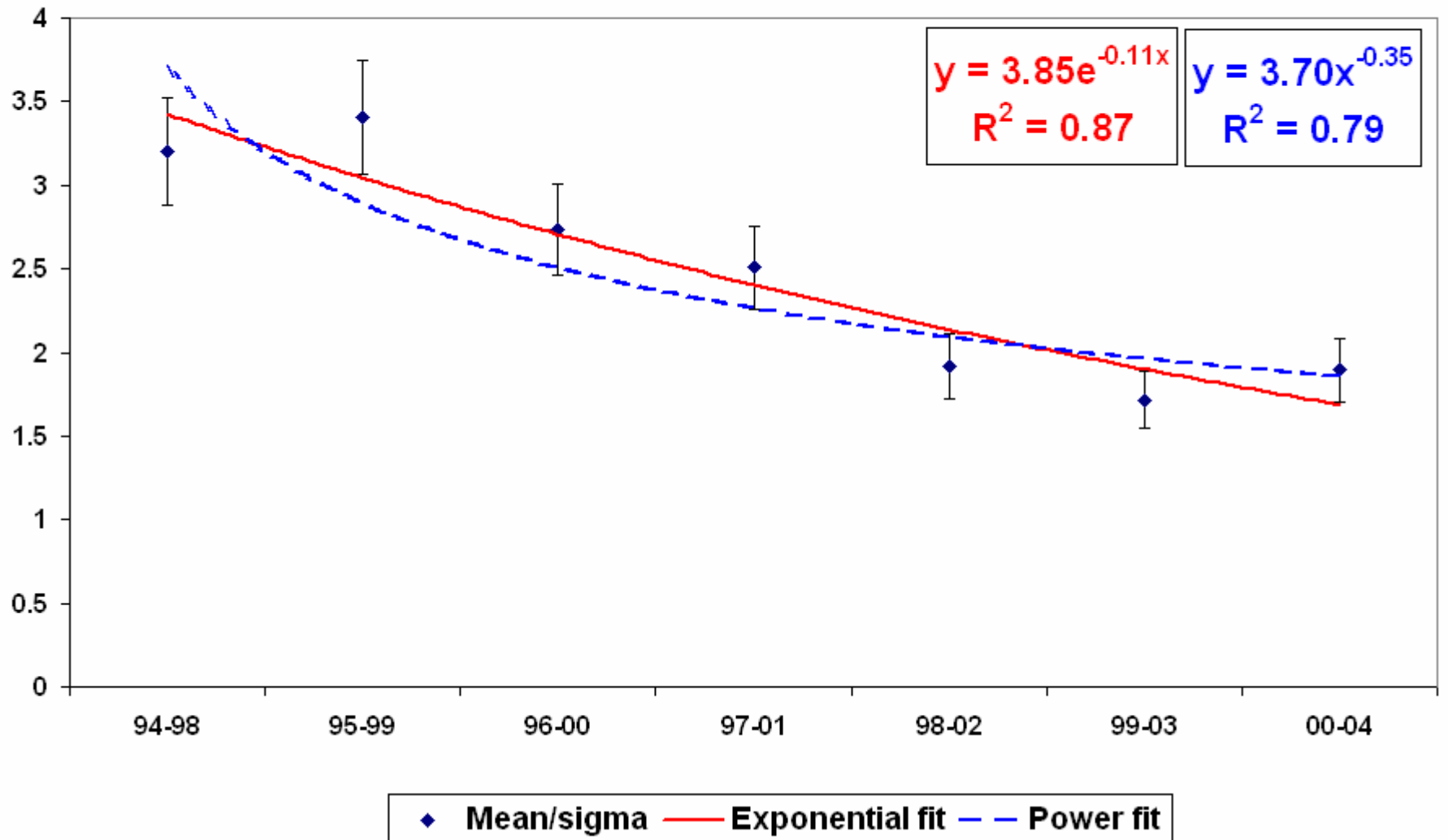
Indicator: GDP/capita. Time window size: 10 years



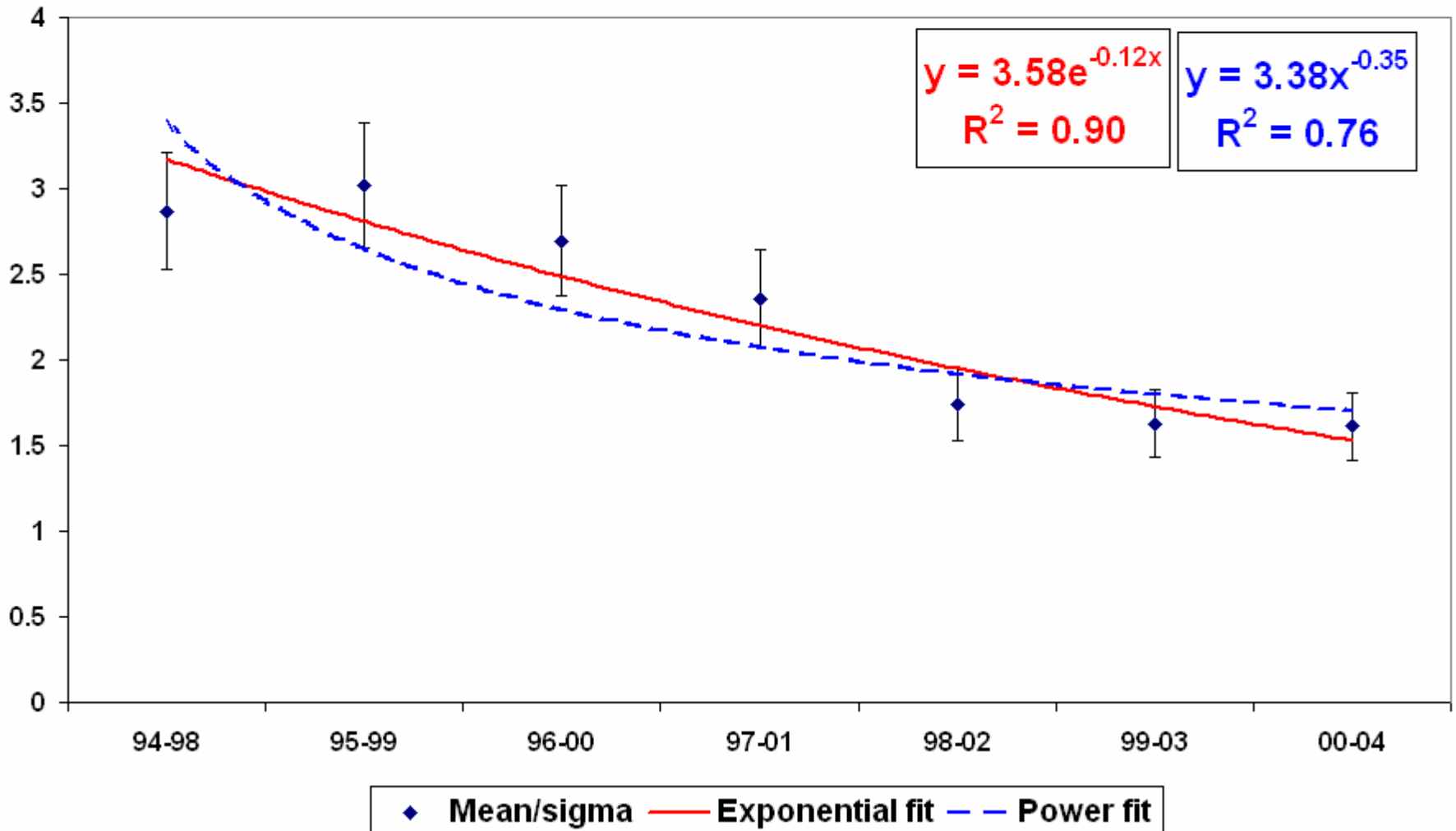
RESULTS (2)

Evidence of globalization effects
in moving 5 Yr time-window size

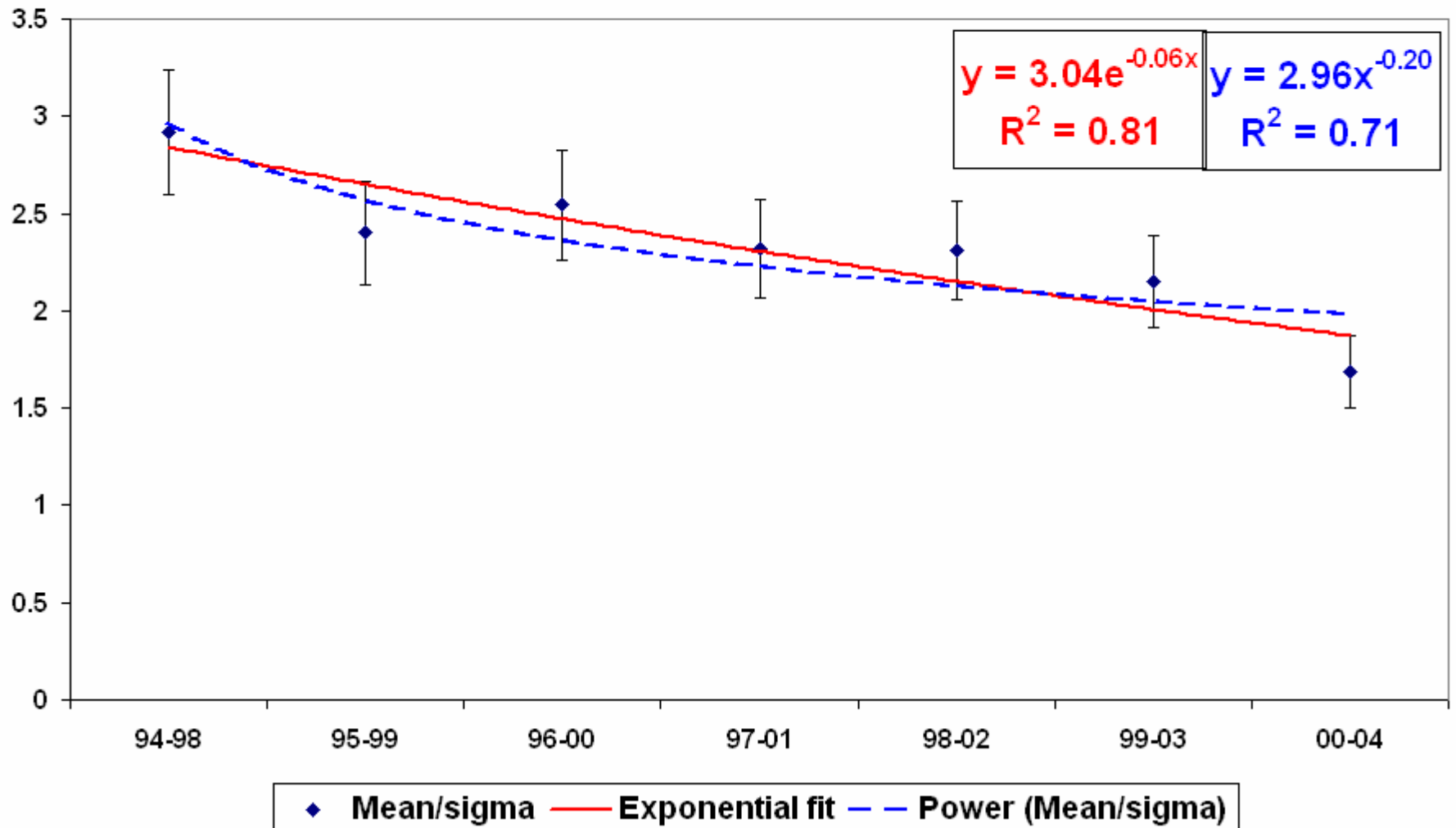
The time evolution of mean statistical distance between EU-15.
Indicator: GDP/capita. Time Window Size: 5 years



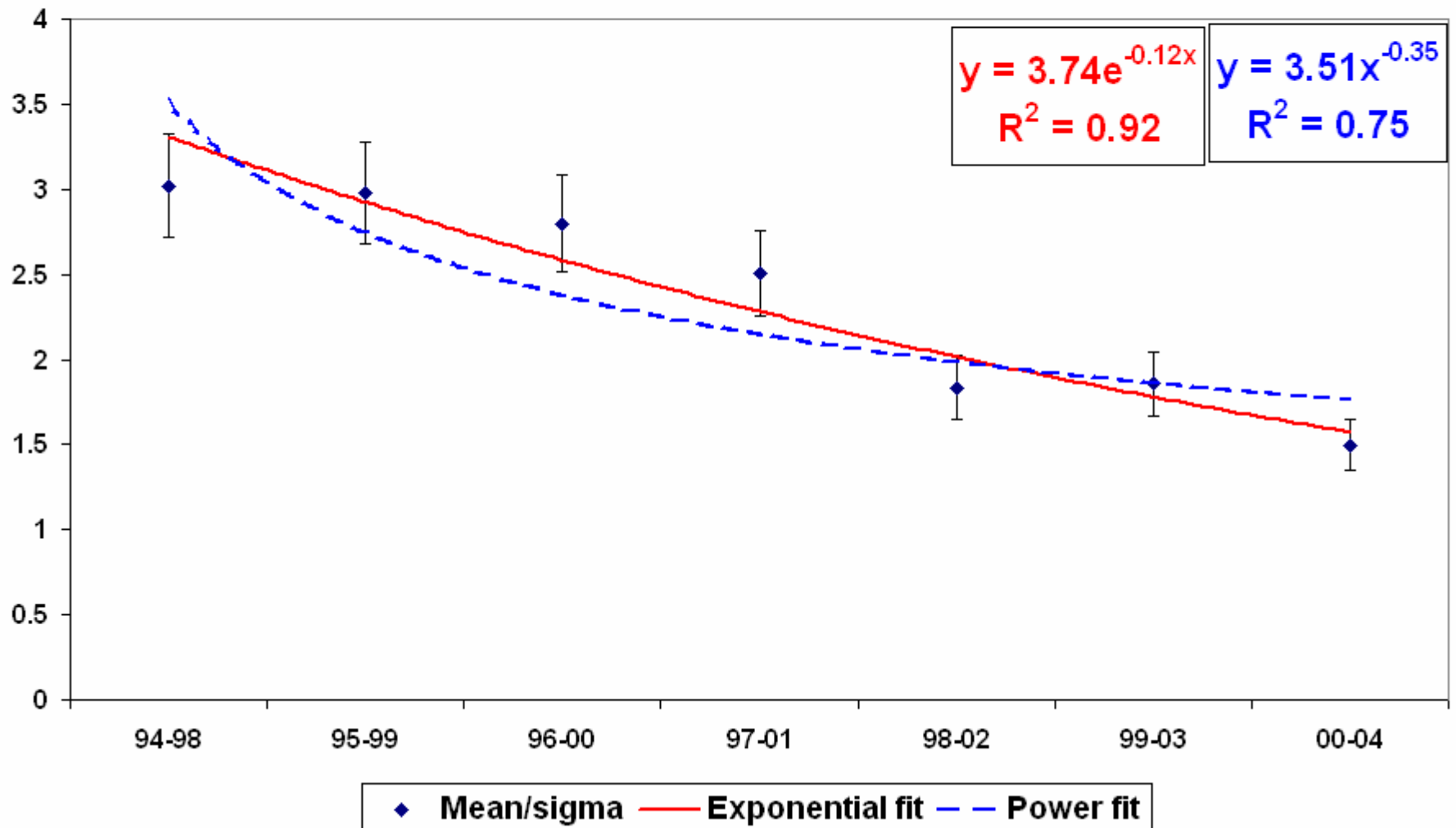
The time evolution of mean statistical distance between EU-15.
Indicator: GDP. Time Window Size: 5 years



The time evolution of the mean statistical distance between EU-15.
Indicator: Final Consumption. Time Window Size: 5 years



The time evolution of the mean statistical distance between EU-15.
Indicator: Gross Capital Growth. Time Window Size: 5 years



RESULTS (3)

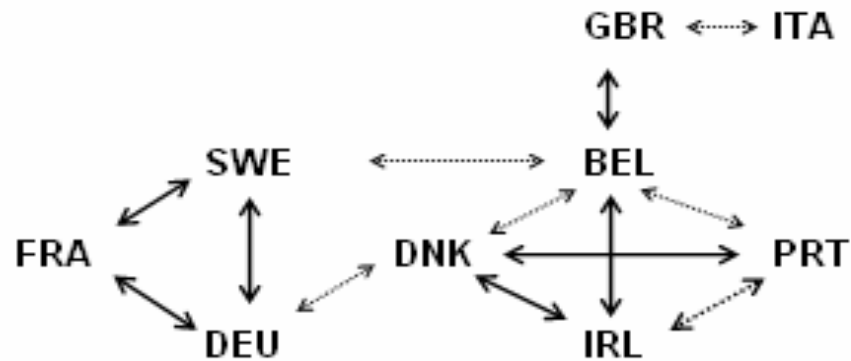
Finding EU-15 country clusters
by MAMLP method

94-98	95-99	96-00	97-01	98-02	99-03	00-04
A	V	E	R	A	G	E
ESP	ESP	LUX	LUX	LUX	DEU	DEU
FIN	NLD	BEL	ESP	DEU	DNK	FRA
GRC	GRC	GBR	NLD	DNK	FRA	NLD
LUX	LUX	GRC	PRT	FRA	NLD	PRT
NLD	FRA	ITA	BEL	GBR	PRT	DNK
AUT	DEU	ESP	DNK	ESP	BEL	AUT
FRA	SWE	DEU	IRL	NLD	LUX	BEL
BEL	AUT	FRA	FIN	PRT	AUT	FIN
DEU	BEL	SWE	GBR	BEL	FIN	GBR
DNK	IRL	IRL	AUT	IRL	IRL	LUX
GBR	DNK	AUT	DEU	SWE	SWE	SWE
IRL	FIN	NLD	FRA	AUT	ESP	ESP
ITA	GBR	PRT	SWE	FIN	GBR	ITA
PRT	PRT	DNK	GRC	ITA	ITA	IRL
SWE	ITA	FIN	ITA	GRC	GRC	GRC

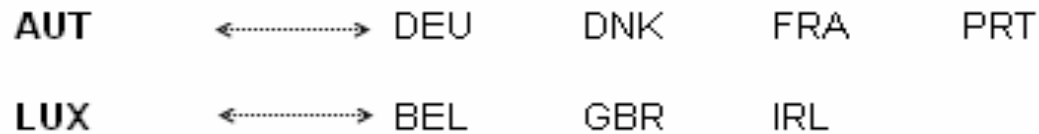
The correlation matrix of country movements inside the hierarchy.
Indicator: GDP. Time window size: 5 years

	AUT	BEL	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL	ITA	LUX	NLD	PRT	SWE
AUT	1.00	0.77	0.88	0.88	0.33	0.69	0.88	0.69	-0.69	0.75	0.71	0.42	0.61	0.89	0.85
BEL		1.00	0.88	0.90	0.41	0.27	0.80	0.94	-0.59	0.92	0.83	0.85	0.23	0.90	0.91
DEU			1.00	0.90	0.61	0.35	0.98	0.86	-0.65	0.85	0.78	0.61	0.52	0.86	0.99
DNK				1.00	0.50	0.58	0.87	0.84	-0.80	0.93	0.67	0.77	0.58	0.99	0.88
ESP					1.00	-0.10	0.61	0.34	-0.38	0.55	0.05	0.36	0.66	0.37	0.64
FIN						1.00	0.42	0.25	-0.62	0.34	0.27	0.14	0.60	0.64	0.26
FRA							1.00	0.79	-0.71	0.81	0.73	0.52	0.60	0.82	0.95
GBR								1.00	-0.52	0.82	0.90	0.85	0.12	0.86	0.86
GRC									1.00	-0.82	-0.38	-0.56	-0.62	-0.76	-0.60
IRL										1.00	0.63	0.85	0.43	0.90	0.87
ITA											1.00	0.59	-0.05	0.73	0.77
LUX												1.00	0.06	0.77	0.65
NLD													1.00	0.50	0.47
PRT														1.00	0.84
SWE															1.00

STRONG CORRELATIONS ($0.9 < C < 1$)



MEDIUM CORRELATIONS ($0.85 < C < 0.9$)



STRONG ANTI-CORRELATIONS ($-1 < C < -0.5$)



The correlation matrix of country movements inside the hierarchy.
 Indicator: Final Consumption Expenditure. Time window size: 5 years

	AUT	BEL	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL	ITA	LUX	NLD	PRT	SWE
AUT	1.00	0.92	1.00	0.23	0.92	0.21	0.38	0.87	0.03	0.92	0.07	-0.34	0.92	0.92	0.92
BEL		1.00	0.94	0.23	1.00	0.45	0.56	0.97	0.28	1.00	0.06	-0.15	1.00	1.00	1.00
DEU			1.00	0.24	0.93	0.24	0.40	0.89	0.07	0.94	0.07	-0.32	0.93	0.93	0.93
DNK				1.00	0.26	0.22	-0.14	0.35	0.75	0.23	-0.41	0.44	0.26	0.26	0.26
ESP					1.00	0.45	0.53	0.97	0.31	1.00	0.04	-0.15	1.00	1.00	1.00
FIN						1.00	0.65	0.49	0.34	0.45	-0.68	0.68	0.45	0.45	0.45
FRA							1.00	0.64	0.05	0.56	-0.05	0.38	0.53	0.53	0.53
GBR								1.00	0.40	0.97	0.03	0.02	0.97	0.97	0.97
GRC									1.00	0.28	-0.11	0.45	0.31	0.31	0.31
IRL										1.00	0.06	-0.15	1.00	1.00	1.00
ITA											1.00	-0.68	0.04	0.04	0.04
LUX												1.00	-0.15	-0.15	-0.15
NLD													1.00	1.00	1.00
PRT														1.00	1.00
SWE															1.00

The correlation matrix of country movements inside the hierarchy.
 Indicator: Gross Capital Formation. Time window size: 5 years

	AUT	BEL	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL	ITA	LUX	NLD	PRT	SWE
AUT	1.00	0.76	0.59	0.68	0.88	0.69	0.88	0.10	0.19	0.45	-0.04	-0.58	-0.12	-0.26	0.94
BEL		1.00	0.47	0.81	0.67	0.79	0.67	0.35	0.15	0.85	-0.02	-0.27	0.32	0.15	0.73
DEU			1.00	0.10	0.64	0.09	0.64	0.05	0.55	0.30	-0.57	-0.02	-0.08	-0.25	0.81
DNK				1.00	0.41	1.00	0.41	0.61	-0.32	0.50	0.56	-0.40	0.24	0.39	0.55
ESP					1.00	0.40	1.00	-0.04	0.61	0.58	-0.35	-0.26	0.11	-0.29	0.83
FIN						1.00	0.40	0.58	-0.37	0.46	0.57	-0.46	0.17	0.35	0.56
FRA							1.00	-0.04	0.61	0.58	-0.35	-0.26	0.11	-0.29	0.83
GBR								1.00	-0.21	0.20	0.63	0.37	0.61	0.91	0.12
GRC									1.00	0.44	-0.76	0.45	0.37	-0.20	0.27
IRL										1.00	-0.26	0.10	0.62	0.21	0.40
ITA											1.00	-0.15	0.12	0.60	-0.21
LUX												1.00	0.73	0.60	-0.46
NLD													1.00	0.78	-0.17
PRT														1.00	-0.27
SWE															1.00

**The correlation matrix of country movements inside the hierarchy.
Indicator: Net Exports. Time window size: 5 years**

	AUT	BEL	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL	ITA	LUX	NLD	PRT	SWE
AUT	1.00	-0.39	0.80	-0.32	0.11	0.02	-0.89	-0.62	0.30	-0.59	0.60	0.02	-0.26	0.84	-0.59
BEL		1.00	-0.65	-0.39	0.09	-0.39	0.15	-0.30	-0.32	0.62	-0.61	-0.39	-0.27	-0.48	0.62
DEU			1.00	-0.07	0.44	-0.05	-0.56	-0.35	0.06	-0.92	0.82	-0.05	0.13	0.93	-0.92
DNK				1.00	0.22	0.85	0.28	0.56	0.58	-0.14	-0.28	0.85	0.86	-0.41	-0.14
ESP					1.00	-0.03	-0.16	-0.37	-0.18	-0.64	0.23	-0.03	0.53	0.30	-0.64
FIN						1.00	-0.13	0.30	0.86	-0.04	-0.29	1.00	0.56	-0.31	-0.04
FRA							1.00	0.82	-0.29	0.47	-0.47	-0.13	0.35	-0.67	0.47
GBR								1.00	0.21	0.34	-0.40	0.30	0.50	-0.57	0.34
GRC									1.00	0.05	-0.35	0.86	0.40	-0.16	0.05
IRL										1.00	-0.82	-0.04	-0.28	-0.81	1.00
ITA											1.00	-0.29	-0.24	0.90	-0.82
LUX												1.00	0.56	-0.31	-0.04
NLD													1.00	-0.25	-0.28
PRT														1.00	-0.81
SWE															1.00

The quadratic sum of correlation coefficients

GDP		Net Exports	
9.08	DNK	5.23	PRT
8.71	PRT	4.92	DEU
8.68	DEU	4.76	IRL
8.47	SWE	4.76	SWE
8.26	IRL	4.41	ITA
8.25	BEL	3.99	AUT
8.21	FRA	3.50	DNK
7.60	AUT	3.24	FRA
7.59	GBR	3.23	FIN
5.68	ITA	3.23	LUX
5.64	GRC	2.91	GBR
5.40	LUX	2.71	BEL
3.25	NLD	2.63	NLD
2.97	ESP	2.49	GRC
2.68	FIN	1.69	ESP

Outlook and conclusion

We found that:

- In large time windows, the “globalization” effect (i.e. the decreasing mean distance between countries) appears as nonlinear and non-monotonic.
- In 5 Yr moving time window size, the process has the attributes of an “aftershock” relaxation: the decreasing trend is well fitted to an exponential.
- The relaxation time of exponential is 8-10 years for GDP, GDP/capita and Gross Capital Growth and larger (~ 16 years) for Final Consumption Expenditure.

We found that:

- The strongly correlated countries in GDP fluctuations can be roughly partitioned into two clusters: **FRA-SWE-DEU** and **IRL-GBR-BEL-DNK-PRT**.
- Six of the mostly correlated countries in GDP fluctuations (SWE, GBR, DEU, BEL, IRL, PRT) appear as strong correlated in the FCE fluctuations, three (FRA, SWE, PRT) – in GCF fluctuations, and two (PRT, SWE) – in Net Export fluctuations.
- The “connectivity” classifications regarding GDP and Net Export fluctuations are quite similar (the “positive correlation” existing between the node degrees)

Further developments

- To apply the Cluster Variation Method considering the countries as *function nodes* and ME indicators as *variable nodes*;
- To describe the clusters by appropriate statistical / thermodynamic quantities;
- To analyse the stability of the clustering structure against removing / addition of one (of more) countries.

$$H = A(2,3)+B(1,2)+De(1,2,4)+Dn(1,3)+E(2,3)+Fi(3,4)+Fr(1,3)+Gb(1,2,3)+Ir(1,2,4)+It(1,4)+L(4)+N(2)+P(1,2,3,4)+S(1,2,3,4)$$

