

SAPIENZA - UNIVERSITÀ DI ROMA

ANNALI DEL DIPARTIMENTO DI METODI
E MODELLI PER L'ECONOMIA,
IL TERRITORIO E LA FINANZA

2014

PÀTRON EDITORE
Bologna 2015

Direttore Responsabile - Director

Alessandra De Rose

Direttore Scientifico - Editor in Chief

Roberta Gemmiti

Comitato Scientifico - Co-editors

Maria Giuseppina Bruno, Francesca Gargiulo, Roberta Gemmiti, Cristina Giudici, Ersilia Incelli, Antonella Leoncini Bartoli, Isabella Santini, Rosa Vaccaro.

Comitato di Redazione - Editorial Staff

Elena Ambrosetti, Maria Caterina Bramati, Filippo Celata, Augusto Frascatani, Maria Rita Scarpitti, Maria Rita Sebastiani, Marco Teodori, Judith Turnbull.

Consulenti Scientifici - Advisory Board

Catherine Wihtol de Wenden (CERI-Sciences Po-CNRS Paris), Raimondo Cagiano de Azevedo (Sapienza - Università di Roma), Maria Antonietta Clerici (Politecnico di Milano), Marina Fuschi (Università di Chieti-Pescara), Alessandra Faggian (The Ohio State University), Alison Brown (Cardiff University), Luciano Pieraccini (Università degli Studi Roma Tre), Silvia Terzi (Università degli Studi Roma Tre), Gennaro Olivieri (Luiss Guido Carli), Giulio Fenicia (Università degli Studi di Bari), Angelo Moioli (Università Cattolica del Sacro Cuore), Filomena Racioppi (Sapienza - Università di Roma); Pablo Koch-Medina (Centro di Finanza e Assicurazioni, Università di Zurigo).

External Reviewers - Blind review

Copyright © 2015 by Pàtron editore - Quarto Inferiore - Bologna

I diritti di traduzione e di adattamento, totale o parziale, con qualsiasi mezzo sono riservati per tutti i Paesi. È vietata la riproduzione parziale, compresa la fotocopia, anche ad uso interno o didattico, non autorizzata.

PÀTRON Editore - Via Badini, 12
Quarto Inferiore, 40057 Granarolo dell'Emilia (BO)
Tel. 051.767003
Fax 051.768252

E-mail: info@patroneditore.com

<http://www.patroneditore.com>

Il catalogo generale è visibile nel sito web. Sono possibili ricerche per autore, titolo, materia e collana. Per ogni volume è presente il sommario, per le novità la copertina dell'opera e una breve descrizione del contenuto.

Stampa: Rabbi s.r.l., Bologna per conto di Pàtron editore.

ISBN: 978-88-555-3310-2

ISSN: 2385-0825

DIMENSION REDUCTION FOR MEASURING THE MULTIDIMENSIONAL DEMOGRAPHIC CONVERGENCE OF THE EUROPEAN UNION POPULATIONS

Summary: The European Union (EU) was founded in 1957 in order to prevent new wars in the Continent. Over time, many institutions and instruments have been introduced for defining and implementing common policies concerning economics, society and politics. All these innovations have contributed to generating the “European people”, beyond the single national peoples, with common institutions and cultural heritage. In fact, each people is also a population, describable in terms of demographic structure as well as in terms of natural and migratory dynamics. Therefore, it is reasonable to ask if, over time, all the national populations of the EU tend to converge towards a European population, with a common demographic profile. Here, a statistical method is proposed for measuring the demographic convergence of the populations of the EU towards a common pattern.

Keyword: demographic convergence, European populations, multiple variability.

1. Introduction

The European Union (EU) – originally the European Economic Community – was founded in 1957 in order to prevent new wars in the Continent. It started as a Customs Union, with the aim of subsequently creating a Common Market for the free circulation of goods, services, people and capitals. Over time, new institu-

* Dipartimento Memotef – Sapienza – Università di Roma.

tions and instruments have been introduced for defining and implementing common policies concerning economics, society and politics. All these innovations have contributed to generating the “European people”, beyond the single national peoples, with common institutions and cultural heritage. In fact, each people is also a population, describable in terms of demographic structure as well as in terms of natural and migratory dynamics.

Therefore, it is reasonable to ask if, over time, all the national populations of the EU tend to converge towards a “European population”, with a common demographic profile. The existence of a European population implies that some common demographic policies could be proposed for dealing with specific problems (such as, for instance, the decline in fertility, the increasingly aging population) that can affect the welfare state.

The aim of this paper is to introduce a statistical method for measuring the demographic convergence of the EU’s populations towards a common pattern in a multidimensional perspective that is considering several variables together at one time. Demographic convergence is a key point of both the First and the Second Demographic Transition theories (FDT and, respectively, SDT). In particular, the FDT theory discusses the decline in birth and mortality rates of the developed populations, which over time would tend towards a steady state with zero growth and a growth in aging (Thompson, 1929; Landry, 1934; Notestein, 1945). The SDT theory is relatively recent and includes other variables (Les-thaeghe, Surkyn, 2004; Van De Kaa, 1987, 2003). It has been introduced for explaining the sharp decline in fertility under the replacement rate, observed in all the populations of industrialized countries (Europe, North America, Australia, New Zealand and Japan), in conjunction with a great transformation in the processes of family’ formation and duration, and the non-occurrence of a stationary population. Compared to the FDT theory, the SDT also studies the role of survival and of international migration in the process of a population’s growth, as well as the implications of aging, of immigration and of the weakness of family structures for the welfare state systems.

Many researchers have already empirically tested demographic convergence using some statistical indices of variability (σ -convergence approach; see, for instance Dorius, 2008; Neumayer, 2004; Sebastiani, 2010). If the variability is low, it means that

the populations are similar to each other and therefore convergence is achieved. Conversely, if the variability is high, the populations are different from each other and therefore convergence is not attained. Specifically, in Sebastiani (2010) an absolute index of variability and the corresponding normalized one (represented respectively by s_{jj} and $s_{jj, norm}$) have been proposed, since they are considered the most adequate indices when the variable is a demographic rate (for instance, crude birth rate or crude death rate). However, almost all of the cited studies have only focused on a one-dimensional perspective, testing separately the convergence of one variable at a time. In fact, each population is characterized by all the different demographic phenomena, so it could be interesting to measure the convergence considering all the variables together. Indeed, populations that are similar to each other with respect to some variables (for instance, in terms of mortality), could differ with respect to others (for instance, in terms of births), indicating that they are in different stages of transition.

Let us consider a set of k variables each observed on n populations. Recently, some statistical indices were introduced for measuring the multiple variability of the n individual k -variate profiles with respect to the average k -variate profile (M.R. Sebastiani, 2014). In particular, because both the FDT and SDT theories concern the changes of births and deaths over time, and the consequent changes of the age structure, the crude birth rates, crude death rates, infant mortality rates and aging index were considered (represented respectively by CBR, CDR, IMR, AI). Since CBR, CDR, IMR and AI are statistical rates, the covariance matrix \mathbf{S} was appropriately defined. Specifically, the variance had the expression of s_{jj} , whereas the covariance was defined similarly to s_{jj} . Moreover, since the rates can differ from each other with respect to their magnitude as well as with respect to their degree of variability, it seemed adequate to consider the correlation matrix \mathbf{R} derived from \mathbf{S} too. For measuring the variability of the n k -dimensional populations, the trace and the determinant of \mathbf{S} and also the determinant of \mathbf{R} were proposed as absolute indices of multiple variability (by symbols, $tr(\mathbf{S})$, $det(\mathbf{S})$ and $det(\mathbf{R})$). Specifically, $tr(\mathbf{S})$ measures the distance of the n populations from the average profile, if the variables are independent of each other; both $det(\mathbf{S})$ and $det(\mathbf{R})$ measure the volume of the k -dimensional space where the n populations are represented, and are more appropriate than

$tr(\mathbf{S})$ when the variables are correlated to each other. Afterwards, with the aim of evaluating how high the degree of convergence is, a linear normalization procedure was applied to each absolute index of multiple variability, after determining for each of them both the minimum and maximum values. So the corresponding normalized indices were obtained, each with values comprised between 0 and 1. Specifically, values near to 0 mean that there is no variability and therefore demographic convergence is achieved. Conversely, values near to 1 indicate that variability is maximum and therefore the individual profiles diverge from each other and from the common one. The normalized indices were applied for measuring the convergence of the populations of the EU. However, that procedure has revealed some drawbacks. Firstly, CBR and CDR might not be the most appropriate variables for studying reproduction and survival, since they are influenced by the age structure of the population. Moreover, if the variables were correlated to each other, $det(\mathbf{S})$ and $det(\mathbf{R})$ could take low values with k increasing, making it difficult to understand if the indices of multiple variability take low values due to convergence or to correlation among variables.

Here, some statistical developments have been introduced with the aim to overcome these drawbacks. Firstly, in addition to CBR, CDR, IMR and AI, for measuring reproduction and the survival of populations have been also used the total fertility rate and the life expectancy at birth by gender (represented respectively by TFR; LEM, for males; LEF, for females). In fact, TFR, LEM and LEF do not depend on the age structure of populations as instead happens with CBR and CDR. However, in general CBR and CDR have been used too, because they could give information on reproduction and survival when data on TFR, LEM and LEF are not yet available. Secondly, the well known statistical technique of the principal component analysis (*pca*) has been applied to the original dataset in order to define new uncorrelated variables (the principal component variables or *pcvs* (Mardia et al., 1979)), on which to measure the multidimensional demographic convergence by means of the normalized indices of multiple variability. By doing this, it can be ensured that the indices will take small values if and only if variability is low, indicating that convergence is achieved. Six different frameworks of the EU have been considered, according to the successive expansions (specifically, with 6,

10, 12, 15, 25 and 27 members) and the six corresponding groups of national populations (represented by the symbols EU6, EU10, EU12, EU15, EU25 and EU27). The multiple variability has been measured within each group over time: a decreasing trend means progressive demographic convergence, whereas stationarity or an increasing trend indicate respectively delays or reversals in the convergence. By comparing the values and trends that a normalized index assumes within the different groups, the most and the least converging groups can be identified and, moreover, the time when lags eventually occur among the processes of convergence of the different groups. Although the proposed method is described with reference to data on the populations of the EU, it can of course be applied more generally to other groups of populations with suitable modifications.

The paper is organized as follows: in Section 2, the data are presented and in Section 3, is described the proposed methodology. Specifically, the matrix of the k variables observed on the n populations, the covariance matrix \mathbf{S} and the correlation matrix \mathbf{R} are defined. Afterwards, it is explained how to use the *pca* and the *pcvs* for testing demographic convergence. In Section 4, are discussed the results obtained by applying the normalized indices of multiple variability for measuring the convergence of the EU's populations. Finally, Section 5 contains some concluding remarks.

2. Data

For each of the EU's 27 populations the yearly series of CBR, TFR, CDR, LEM, LEF, IMR and AI have been taken, for all the years from 1989 to 2010¹ (source: Eurostat). Each yearly series describes an individual demographic profile. For each variable, the individual profiles of populations belonging to the first four groups (that comprise the Northern, Western and Southern populations) are quite similar to each other. Conversely, the profiles of the Eastern populations and those of the island populations (those of Malta and Cyprus), often differ from each other as well as when compared with those of the other European populations.

¹ Source EUROSTAT: http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/main_tables.

In particular, in the case of CDR, AI, LEM and LEF, the differences among profiles that occurred during the 1990s are due to a worsening in living conditions after the fall of the Berlin Wall and the collapse of the Communist regimes. Instead, in the case of CBR and TFR, the differences among profiles depend on the fact that in the Eastern and in the island populations fertility declines later than in the Northern and Western ones, just as the SDT theory states. However, it is very likely that in general the individual profiles of the Eastern and of the island populations are going to become similar to those of the other European populations.

For each variable, the six European common profiles have been determined, one for each group. Specifically, in the case of CBR, CDR, IMR and AI, for every year the number of the demographic occurrences in the group (namely, births, deaths or number of elderly people) has been divided by the overall reference population. Instead, in the case of TFR, LEM and LEF, for every year the value corresponding to the overall population of the group has been approximated by averaging the values that the variable takes over the n populations of the group. For each variable, except for TFR, the six common profiles tend to be adjacent or even to overlap each other. The common profiles of CBR, CDR and IMR tend to decrease (specifically, CBR from 11.6-12.5 per 1,000 to about 10.1-10.8 per 1,000; CDR from 10.2-10.4 per 1,000 to 9.3-9.7 per 1,000; IMR from 7.7-10.7 per 1,000 to 3.5-4 per 1,000). Instead, the common profiles of LEM, LEF and AI increase considerably (respectively, LEM from 70.6-72.6 to 75.6-78.3; LEF from 77.5-79.3 to 81.9-83.7; AI from 69-81.2 per 100 to 111.7-123 per 100). The common profiles of TFR calculated for EU6, EU10, EU12 and EU15 tend to increase slightly from 1.53-1.62 to 1.68-1.7, due to rising fertility in the populations of Western and Northern Europe. Instead, the common profiles of TFR calculated for EU25 and EU27 tend to decrease from 1.77-1.79 to about 1.6, in cause of declining fertility in the Eastern populations and in Malta and Cyprus.

In Sebastiani (2010), the results obtained by applying $s_{jj, norm}$ to CBR, CDR, IMR and AI for testing the one-dimensional convergence of the EU's populations had been showed. These populations converge towards a common pattern, which is typical of a "mature" population with low natural increase and a high aging rate. Indeed, for each variable and within each group, variability is always low; moreover, for CDR and IMR, it decreases over time.

Conversely, in the case of CBR and AI (but also for TFR, LEM, LEF), it has risen slightly over the last few years, meaning that some weak reversals in convergence have occurred. Specifically, for CBR and TFR this is probably due to the relatively high increase of births in some very large populations (for instance, in the French population because of the effect of some demographic policies as well as of the higher fertility of immigrant women). Instead, in the case of LEM, LEF, and AI the progressive rise in variability is due to an abrupt worsening of the living conditions of some Eastern populations (specifically, in the Baltic States, Romania and Bulgaria) compared to the other populations.

3. Methodology

Let us use a general notation. Let $S = \{1, \dots, h, \dots, n\}$ and $V = \{1, \dots, j, \dots, k\}$ be two sets of labels representing, respectively, n populations belonging to the same group and k variables (where n and k are fixed integer numbers). Specifically, here $n = 6, 10, 12, 15, 25, 27$ according to the group considered and $k = 7$.

Let $\mathbf{Y} = \{\bar{y}_{hj}\} \in \mathcal{M}_{n,k}$ be the matrix of the k variables observed in the n populations. Specifically, since CBR, CDR, IMR and AI are statistical rates, we have: $y_{hj} = x_{hj}/p_{hj}$, where x_{hj} and p_{hj} represent, respectively, the number of the demographic events (for instance, births, deaths or elderly people) and the reference population for x_{hj} . Let $\bar{\mathbf{Y}} = \{\bar{y}_{hj}\} \in \mathcal{M}_{n,k}$ be the average matrix of the variables, where $\bar{y}_{hj} = \bar{y}_j$ is the mean of the n values y_{hj} . In particular, for CBR, CDR, IMR and AI, it has been defined:

$$\bar{y}_j = \frac{\sum_{h=1}^n y_{hj} p_{hj}}{\sum_{h=1}^n p_{hj}} = \frac{\sum_{h=1}^n x_{hj}}{\sum_{h=1}^n p_{hj}}, \text{ whereas for the other variables } \bar{y}_j \text{ has the}$$

usual arithmetic mean's formula.

Let $\mathbf{S} = \{s_{ij}\} \in \mathcal{M}_{k,k}$ be the covariance matrix. For each variable that is a statistical rate, since its mean is the weighted average of the n rates y_{hj} , it could be appropriate take the variance and the covariance as weighted averages as well. Specifically, for the variance it has been defined:

- if j is CBR, CDR, IMR or AI:

$$s_{jj} = \frac{\sum_{h=1}^n (y_{hj} - \bar{y}_j)^2 \frac{P_{hj}}{\sum_{h=1}^n P_{hj}}}{\sum_{h=1}^n P_{hj}} \quad (1)$$

where s_{jj} is the same index already introduced in Sebastiani (2010);

- if j is TFR, LEM or LEF:

$$s_{jj} = \frac{\sum_{h=1}^n (y_{hj} - \bar{y}_j)^2}{n}. \quad (2)$$

For the covariance, it has been defined:

- if i and j are CBR, CDR, IMR or AI ($i \neq j$):

$$s_{ij} = \frac{\sum_{h=1}^n (y_{hi} - \bar{y}_i)(y_{hj} - \bar{y}_j)}{\sqrt{\sum_{h=1}^n P_{hi} \sum_{h=1}^n P_{hj}}} \quad (3)$$

- if i and j are TFR, LEM or LEF ($i \neq j$):

$$s_{ij} = \frac{\sum_{h=1}^n (y_{hi} - \bar{y}_i)(y_{hj} - \bar{y}_j)}{n}; \quad (4)$$

- if i is TFR, LEM or LEF and j is CBR, CDR, IMR or AI:

$$s_{ij} = \frac{\sum_{h=1}^n (y_{hi} - \bar{y}_i)(y_{hj} - \bar{y}_j)}{\sqrt{\frac{1}{n} \sum_{h=1}^n P_{hj}}} \quad (5)$$

where s_{ij} is the same index already introduced in Sebastiani (2014).

Let $\mathbf{R} = \{r_{ij}\} \in \mathcal{M}_{k,k}$ be the correlation matrix; obviously, it is:

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}.$$

The *pca* has been applied to the matrix $(\mathbf{Y} - \bar{\mathbf{Y}})$ in order to obtain q_s *pcvs* ($q_s \leq k$), which are uncorrelated to each other, and are to be used for testing demographic convergence. Specifically, the covariance matrix $\mathbf{S}^* = \{s_{ij}^*\} \in \mathcal{M}_{q_s, q_s}$ of the *pcvs* has been considered and the multiple variability of the *pcvs* has been summarized by means of $tr(\mathbf{S}^*)$ and $det(\mathbf{S}^*)$. Remembering that the j 'th *pcv* is the eigenvector corresponding to the j 'th greatest eigenvalue of \mathbf{S} (symbols ${}_s \lambda_{j'}$) and, moreover, that its variance s_{jj}^* is equal to ${}_s \lambda_{j'}$ ($j' = 1, \dots, q_s$), we have: $tr(\mathbf{S}^*) = \sum_{j'=1}^{q_s} {}_s \lambda_{j'}$ and $det(\mathbf{S}^*) = \prod_{j'=1}^{q_s} {}_s \lambda_{j'}$.

Since the original variables show different degrees of magnitude and scale and, moreover, are expressed by means of different units of measure, it is appropriate to believe that applying the *pca* to the standardized matrix $((\mathbf{Y} - \bar{\mathbf{Y}})\mathbf{D})$ too, where

$\mathbf{D} = diag\left(\frac{1}{\sqrt{s_{11}}}, \dots, \frac{1}{\sqrt{s_{jj}}}, \dots, \frac{1}{\sqrt{s_{kk}}}\right)$, is an appropriate procedure. In this

case, the *pcvs* are the eigenvectors of the matrix \mathbf{R} and the variance s_{jj}^{**} of the j 'th *pcv* ($j' = 1, \dots, q_R \leq k$) is equal to the j 'th greatest eigenvalue of \mathbf{R} (symbol ${}_R \lambda_{j'}$). Considering the covariance matrix $\mathbf{S}^{**} = \{s_{ij}^{**}\} \in \mathcal{M}_{q_R, q_R}$ of the q_R *pcvs*, $det(\mathbf{S}^{**})$ has been taken as absolute index of multiple variability. Thus we have: $det(\mathbf{S}^{**}) = \prod_{j'=1}^{q_R} {}_R \lambda_{j'}$.

With the aim of evaluating how high the degree of convergence is, a linear normalization procedure has been applied to each absolute index of multiple variability, after determining for each of them both the minimum and maximum values. So the corresponding normalized indices $tr(\mathbf{S}^*)_{norm}$, $det(\mathbf{S}^*)_{norm}$ and $det(\mathbf{S}^{**})_{norm}$ have been obtained, all with values comprised between 0 and 1. In order to determine numerically the values of all the indices (absolute and normalized), an algorithm in R language (R Development Core Team, 2012) has been implemented. In particular, since the matrices \mathbf{S} and \mathbf{R} are positively or semi-positively defined, we have:

$$0 \leq tr(\mathbf{S}^*) = \sum_{j'=1}^{q_s} {}_s \lambda_{j'} \leq \sum_{j=1}^k \max(s_{jj}) \Rightarrow 0 \leq tr(\mathbf{S}^*)_{norm} = \frac{\sum_{j'=1}^{q_s} {}_s \lambda_{j'}}{\sum_{j=1}^k \max(s_{jj})} \leq 1 \quad (6)$$

$$0 \leq \det(\mathbf{S}^*) = \prod_{j'=1}^{q_S} \lambda_{j'} \leq \left(\frac{\sum_{j=1}^k \max(s_{jj})}{q_S} \right)^{q_S} \Rightarrow 0 \leq \det(\mathbf{S}^*)_{norm} = \frac{\prod_{j'=1}^{q_S} \lambda_{j'}}{\left(\frac{\sum_{j=1}^k \max(s_{jj})}{q_S} \right)^{q_S}} \leq 1 \quad (7)$$

$$0 \leq \det(\mathbf{S}^{**}) = \prod_{j'=1}^{q_R} \lambda_{j'} \leq \left(\frac{k}{q_R} \right)^{q_R} \Rightarrow 0 \leq \det(\mathbf{S}^{**})_{norm} = \frac{\prod_{j'=1}^{q_R} \lambda_{j'}}{\left(\frac{k}{q_R} \right)^{q_R}} \leq 1 \quad (8)$$

where $\max(s_{jj})$ represents the maximum value of s_{jj} ($j = 1, \dots, k$).

4. Measuring the multidimensional demographic convergence of the EU's populations

For each group of populations and for each year between 1989 and 2010, the matrices \mathbf{S} and \mathbf{R} have been constructed, $tr(\mathbf{S}^*)_{norm}$, $det(\mathbf{S}^*)_{norm}$ and $det(\mathbf{S}^{**})_{norm}$ have been calculated and the pattern of these indices over time has been studied. The choice of q_S and q_R , namely the number of *pcvs* to be extracted respectively from \mathbf{S} and from \mathbf{R} , is crucial. In both cases, it has been decided to consider the eigenvectors corresponding to the highest eigenvalues of \mathbf{S} (or of \mathbf{R}), so they allowed us to account for at least 75 per 100 of the total variance of the matrix $(\mathbf{Y} - \bar{\mathbf{Y}})$ (or of the matrix $((\mathbf{Y} - \bar{\mathbf{Y}})\mathbf{D})$, as it is the case).

Tables 1 and 2 show, separately for each group of populations, the summary statistics (calculated over the period 1989-2010) of the percentage of the total variance extracted respectively from the matrices $(\mathbf{Y} - \bar{\mathbf{Y}})$ and $((\mathbf{Y} - \bar{\mathbf{Y}})\mathbf{D})$. In terms of the reduction of dimensionality, apparently the *pca* performs better for $(\mathbf{Y} - \bar{\mathbf{Y}})$ than for $((\mathbf{Y} - \bar{\mathbf{Y}})\mathbf{D})$. Indeed, for each group of populations and for each year, by applying the *pca* to the first matrix it was necessary to keep only 1 or 2 *pcvs* for explaining at least 75 per 100 of

the total variance of the original variables. In particular, the percentage values obtained for the groups EU25 and EU27 are significantly higher than those obtained for the other groups. Instead, by applying the *pca* to the matrix $((\mathbf{Y} - \bar{\mathbf{Y}})\mathbf{D})$, for each group of populations and for each year, it was necessary to keep 2 or 3 *pcvs* in order to reach the threshold value of 75 per 100; the percentage values obtained for the first four groups are significantly higher than those obtained for EU25 and EU27. Moreover, for each group of populations and for each year, the percentage of total variance explained by applying the *pca* to the matrix $(\mathbf{Y} - \bar{\mathbf{Y}})$ is almost always greater than the percentage of variance explained by applying the *pca* to the matrix $((\mathbf{Y} - \bar{\mathbf{Y}})\mathbf{D})$; this fact is particularly evident in case of the groups EU25 and EU27.

Tab. 1 - Percentage of the total variance extracted from the matrix $(\mathbf{Y} - \bar{\mathbf{Y}})$ according to the group of populations (summary statistics for the period 1989-2010).

<i>Summary statistics</i>	<i>EU6</i>	<i>EU10</i>	<i>EU12</i>	<i>EU15</i>	<i>EU25</i>	<i>EU27</i>
Minimum	75.27	75.05	75.53	75.21	95.82	95.10
Maximum	96.24	96.98	94.67	94.89	97.94	97.48
Mean	86.88	81.79	81.05	83.78	96.78	96.32
Median	88.35	80.75	80.55	80.88	96.78	96.44

Source: elaborations on EUROSTAT data.

Tab. 2 - Percentage of the total variance extracted from the matrix $((\mathbf{Y} - \bar{\mathbf{Y}})\mathbf{D})$ according to the group of populations (summary statistics for the period 1989-2010).

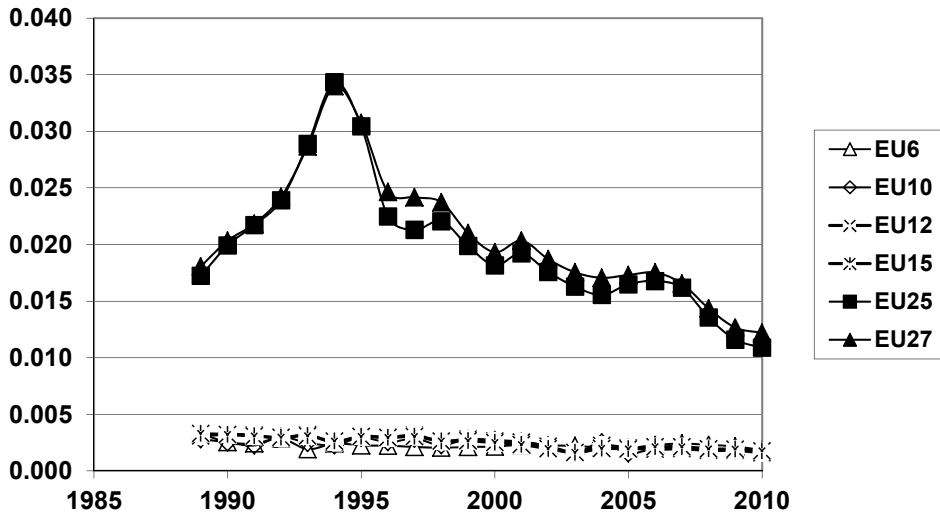
<i>Summary statistics</i>	<i>EU6</i>	<i>EU10</i>	<i>EU12</i>	<i>EU15</i>	<i>EU25</i>	<i>EU27</i>
Minimum	79.05	75.03	76.26	75.10	75.15	75.61
Maximum	91.58	90.57	91.25	87.24	89.53	89.74
Mean	84.69	82.13	82.65	83.38	80.70	81.00
Median	85.03	80.40	84.12	84.43	80.15	80.93

Source: elaborations on EUROSTAT data.

Figures 1 and 2 show, respectively, the pattern of $tr(\mathbf{S}^*)_{norm}$ and of $det(\mathbf{S}^*)_{norm}$ over time, according to the group of populations considered. The results indicate that European populations converge

towards a common pattern: indeed, for each EU-group, the values of both the indices are always very near to zero (on average, about 1-2 per 1,000 for the first four groups and about 20 per 1,000 for EU25 and EU27). Moreover, in all groups, on average both the indices decrease over time (on average, the (compound) variation is between -3 per 100 and -2 per 100). In particular, the values obtained by calculating $det(\mathbf{S}^*)_{norm}$ are almost identical to those obtained by means of $tr(\mathbf{S}^*)_{norm}$. This is due to the fact that the $q_s pcvs$ are independent of each other and, as a consequence, $tr(\mathbf{S}^*)$ and $det(\mathbf{S}^*)$ are equivalent, and so it is for the corresponding normalized indices too. Even if the values of the normalized indices are negligible, it can be observed that variability within EU25 and within EU27 is slightly higher than within the other groups. Moreover, it increases slightly up to 1994, whereas in the first four groups it is stationary during all period. This is due to the fact that EU25 and EU27 also comprise the populations of Eastern Europe and those of the Mediterranean islands, which are at a younger stage of demographic transition than the other populations of the first four groups that instead have already experienced the SDT. In order to highlight the difference among the populations of the first four groups and the other populations belonging to the EU, for each variable has been calculated the (mean) Euclidean distance among the common profile of EU15 and the individual profiles of

Fig. 1 - $tr(\mathbf{S}^*)_{norm}$ according to the group of populations.



the other populations that belong respectively to the groups EU25 and EU27 (Fig. 3 (a), (b), (c)). For CDR and LEM, the distance is increasing over time in cause of the worsening in living conditions in some Eastern populations (namely, in Romania, Bulgaria and the Baltic States), that has caused a considerable rise in mortality and a reduction of the life expectancy at birth above all for males. For LEF, the distance is quite stable over time. Instead, for all the other variables the distance is decreasing in spite of some fluctuations. In particular, for TFR the distance has begun to reduce considerably since the 90s, due to the sharp fall in both the Cypriot and Maltese populations and also in some Eastern ones (namely, those of Poland, Romania and Slovakia). However, from 1994 to about 2004-2006 the distance is slightly increased, because of the rising TFR for common profile of EU15, in contrast with the continuous decline in the other populations of EU25 and EU27.

Figure 4 shows the pattern of $det(\mathbf{S}^{**})_{norm}$ over time. The results differ markedly from those obtained by means of the other two indices. Indeed, in all the groups of populations, variability is quite different from zero (on average, values are equal, respectively, to about 0.5 for EU10, EU12 and EU15, and to 0.6 for EU6, EU25 and EU27). This means that the individual profiles often differ notably from the common profile and also from each

Fig. 2 - $det(\mathbf{S}^*)_{norm}$ according to the group of populations.

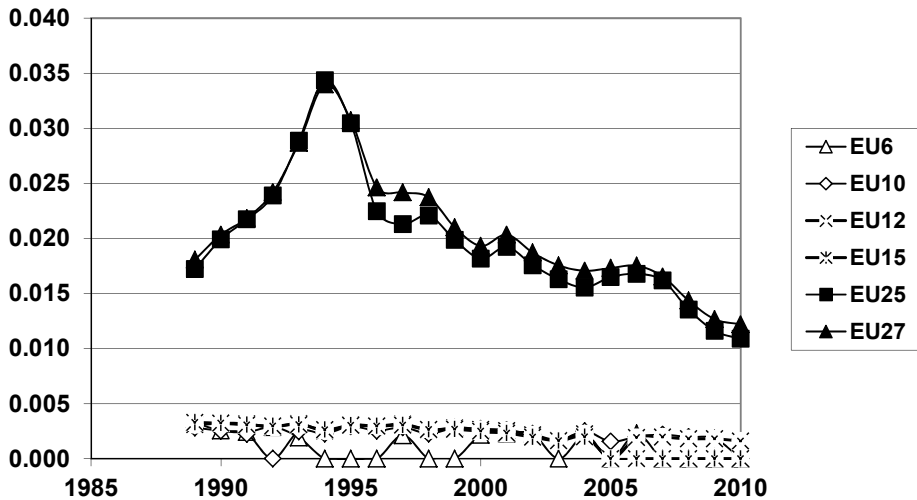
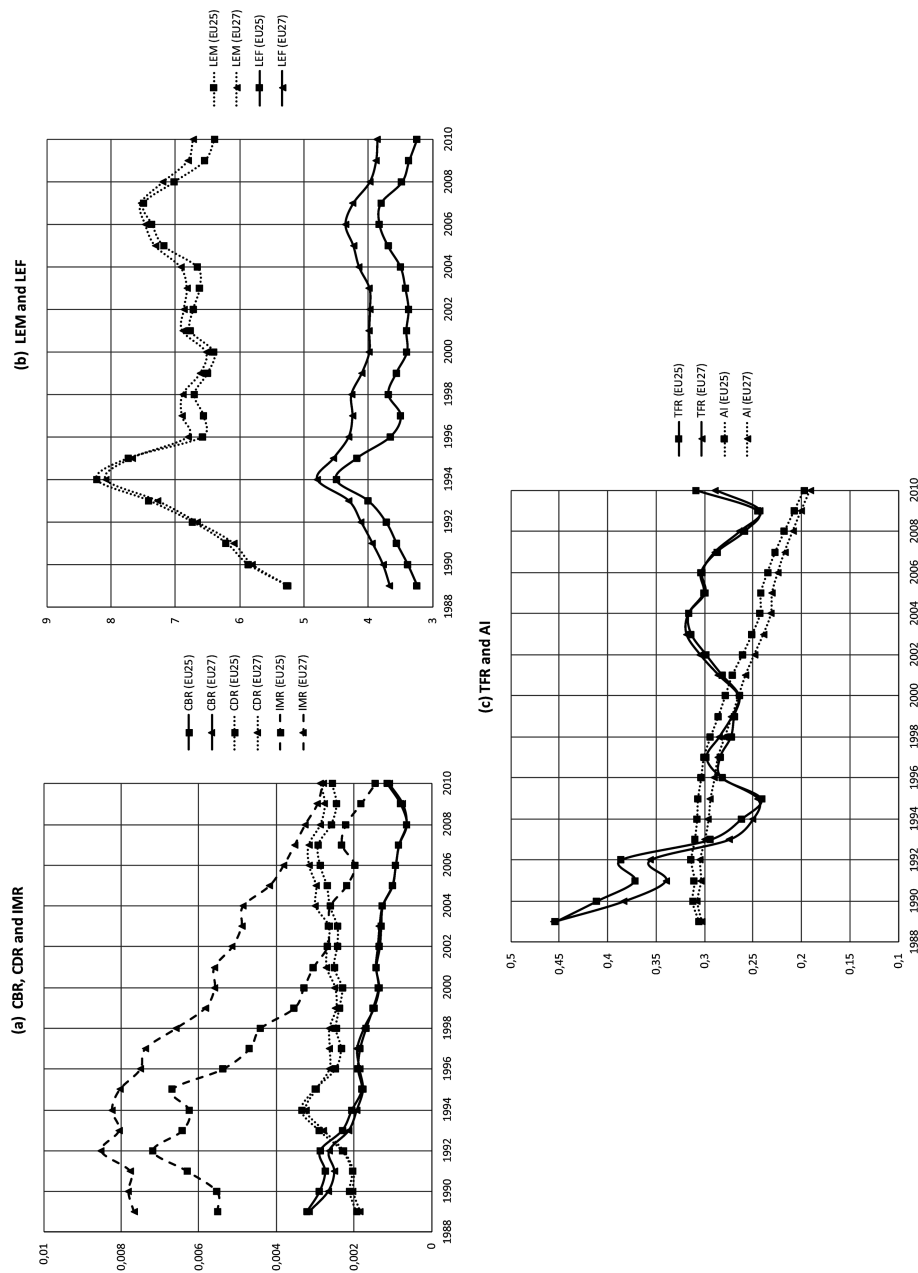
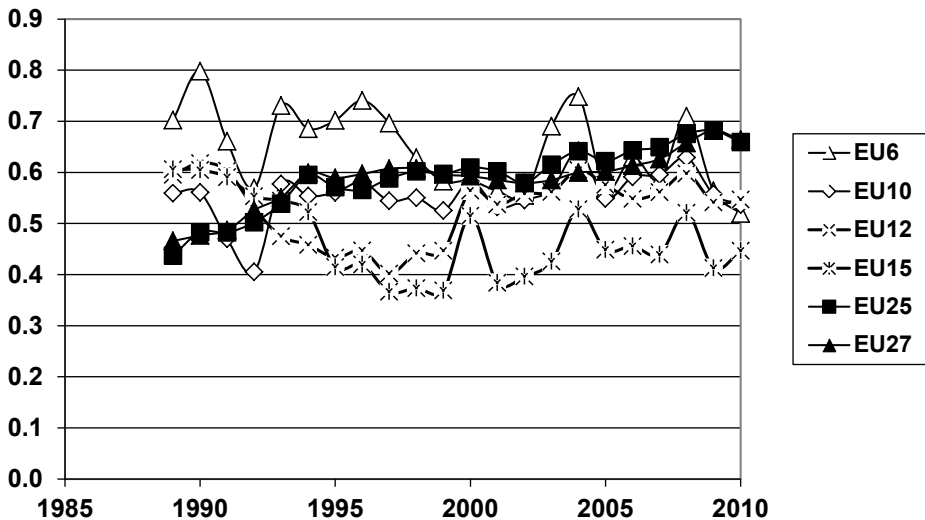


Fig. 3 (a), (b), (c) - (Mean) Euclidean distance among the common profile of EU15 and the individual profiles of the other populations of the groups EU25 and EU27, according to the variable and to the group of populations.



other, and therefore there is no demographic convergence. The most heterogeneous groups are EU25 and EU27 (as already observed by means of both $tr(\mathbf{S}^*)_{norm}$ and $det(\mathbf{S}^*)_{norm}$), and also EU6. The higher variability within EU6 than within EU10, EU12 and EU15, probably depends on the fact that the differences between the German and Italian populations (more elderly people and with lowest fertility) with respect to the other populations of the first four EU-groups, are much more evident in the smaller group (namely, EU6) than in the larger ones. The six curves that represent $det(\mathbf{S}^{**})_{norm}$ show different trends from each other, indicating that there are evident discrepancies and lags between the processes of convergence of the groups. In fact, for the first four groups, the curves display several reversals and delays; moreover, variability seems to be stationary or at the most slightly decreasing over time (on average, the (compound) variation is comprised between -1.4 per 100 and -0.4 per 100). Instead, within EU25 and EU27, variability is constantly increasing (on average, the (compound) variation is equal to 2 per 100 for EU25 and to 1.7 per 100 for EU27), indicating that demographic convergence is still far away.

Fig. 4 $det(\mathbf{S}^{**})_{norm}$ according to the group of populations.



5. Concluding remarks

In this paper, a statistical method is proposed for testing the demographic convergence of the EU's populations towards a common pattern. Compared to the most traditional studies existing in the literature, which measure the demographic convergence of one variable at a time, a multidimensional perspective is adopted that consider all variables together at one time, thus aiming to go further and integrate the partial results obtained by the one-dimensional studies. This is part of the progress made with respect to the traditional methods. Another novelty compared to the traditional methods is that here it is proposed to use some normalized indices of multiple variability, which only take values comprised between 0 and 1. By applying them, a precise assessment can be made as to whether or not demographic convergence is achieved and, moreover, how high the degree of the similarity is among the individual profiles. The method has been implemented for testing the convergence of the populations of the EU from 1989 to 2010, considering six different groups of populations corresponding to the successive expansions of the EU.

The results obtained by means of $tr(\mathbf{S}^*)_{norm}$ and $det(\mathbf{S}^*)_{norm}$, show that European populations converge towards a common pattern and, moreover, that within EU25 and EU27 the convergence is weaker and slower than within the other groups. In fact, EU25 and EU27 also comprise the populations of Eastern Europe and those of the Mediterranean islands, which are in a younger stage of demographic transition than the other populations.

More interesting are the results obtained by applying $det(\mathbf{S}^{**})_{norm}$. Indeed, variability is relatively high in all groups of populations, indicating that the individual profiles often differ notably from the common profile and also from each other, and therefore there is no demographic convergence. The most heterogeneous groups are EU25, EU27 and EU6. Moreover, the six curves that represent $det(\mathbf{S}^{**})_{norm}$ show that there are evident discrepancies and lags between the convergence processes of the groups. In particular, whereas within the first four groups variability is almost stationary or at the most slightly decreasing over time, within EU25 and EU27 it is constantly increasing, indicating that demographic convergence is still far away.

Obviously, in general the results can vary according to the

number of *pcvs* to be extracted. In the literature, there are many criteria for determining this number. Here, it has been decided to apply the criterion of the variance explained: for each group of populations and for each year, have been kept enough *pcvs* to explain a significant percentage of the total variance of the original (eventually standardized) variables. Since here the goal in applying *pca* is the parsimony (namely, to explain variance with as few uncorrelated factors as possible), an appropriate threshold value for this percentage can be 75 per 100. In terms of reducing dimension, the results are satisfactory. For each group, it was always necessary to extract a few *pcvs* (1 or 2, in the case of the original data matrix; 2 or 3, in the case of the standardized variables) to explain significant percentages of the total variance (the median – over time – of the percentage of the variance explained is greater than 80 per 100). However, the proposed methodology has been applied again, thus increasing the threshold value for the percentage of the variance to be explained. For instance, at the 90 per 100 level, it was necessary to keep more *pcvs*, as could be expected (from 2 to 4). Moreover, in terms of the normalized indices of multiple variability, the results obtained by means of $tr(\mathbf{S}^*)_{norm}$ and $det(\mathbf{S}^*)_{norm}$ were quite analogous to those obtained when the percentage of the total variance was fixed at 75 per 100. Instead, the results obtained by applying $det(\mathbf{S}^{**})_{norm}$ were quite unstable over time: the six curves that represent this index showed several jumps, corresponding to the year when the number of *pcvs* extracted increases. This probably depends on the fact that, by using more *pcvs*, some sources of spurious variability have been introduced, that disturb the pattern of the normalized index.

Future work will be concerned with some applied developments in order to be able to include other demographic variables too, such as, for instance, those considered in the SDT theory (namely, the migration rate, marriage rate, etcetera). Moreover, taking into account the differences among the stages of demographic transition of the European populations by geographical area (distinguishing among Western, Central, Eastern and Southern Europe), some adequate normalized variability indices will be introduced, that are based on the decomposition of the total variance. Finally, the distributional properties of all the indices proposed will be studied, so that they can also be used for statistical inference.

Acknowledgments: I am very grateful to Professor Raimondo Cagiano de Azevedo for introducing me to the research area of demographic convergence. I would also like to thank Professor Alessandra De Rose for her useful suggestions that led to a significant improvement in the presentation of the paper. Many thanks are also due to the anonymous referee for his constructing comments on an earlier draft of this article.

References

- DORIUS S.F. (2008), Global Demographic Convergence? A Reconsideration of Changing Intercountry Inequality in Fertility, *Population and Development Review*, 34,3, pp. 519-537.
- LESTHAEGHE R., SURKYN J. (2004), *When History moves on: the Foundations and Diffusion of a Second Demographic Transition*, Seminar on Ideational Perspectives on International Family Change, Population Studies Center, Institute for Social Research (ISR), University of Michigan.
- LANDRY A. (1934), *La Révolution Démographique*, Sirey, Paris.
- MARDIA K.V., KENT J.T., BIBBY J.M. (1979), *Multivariate analysis*, Academic Press, London.
- NEUMAYER E. (2004), HIV/AIDS and cross-national convergence in life expectancy, *Population and Development Review*, 30, pp. 727-742.
- NOTESTEIN F. (1945), Population: the long view, in SCHULTZ T. (ed), *Food for the world*, Chicago, University of Chicago Press, pp. 36-57.
- R DEVELOPMENT CORE TEAM (2012), *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Wien, <http://www.R-project.org>.
- SEBASTIANI M.R. (2010), Measuring the demographic convergence of the European populations by a normalized variability index, *Statistica & Applicazioni*, 2, pp. 181-198.
- SEBASTIANI M.R. (2014), Measuring the multidimensional demographic convergence by indices of multiple variability, in *Statistical Methods and Applications in a Historical Perspective. Selected Issues*, Springer, Heidelberg.
- THOMPSON W.S. (1929), Population, *American Journal of Sociology*, 34, pp. 959-975.
- VAN DE KAA D.J. (1987), Europe's second demographic transition, *Population Bulletin*, 42, 1, pp. 1-59.
- VAN DE KAA D.J. (2002), *The Idea of a Second Demographic Transition in Industrialized Countries*, Sixth Welfare Policy Seminar of the National Institute of Population and Social Security, Tokyo, Japan, 29 January 2002.

Riassunto: L'Unione Europea (EU) è stata fondata nel 1957 allo scopo di prevenire l'insorgere di nuove guerre nel Continente. Nel tempo, sono state introdotte molte istituzioni e strumenti con l'obiettivo di definire ed implementare politiche comuni di natura economica, sociale e politica. Tutte queste innovazioni hanno contribuito a generare il "popolo europeo", al di là dei singoli popoli nazionali, con istituzioni e tradizioni culturali comuni. In realtà, ogni popolo è al tempo stesso una popolazione descrivibile in termini di struttura demografica e di dinamiche naturali e migratorie. Di conseguenza, è lecito chiedersi se nel tempo tutte le popolazioni nazionali dell'EU tendano a convergere verso una "popolazione europea" con un profilo demografico comune.

Résumé: L'Union Européenne (UE) a été fondée en 1957 afin d'éviter l'émergence de nouvelles guerres sur le continent. Au fil du temps, de nombreuses institutions ont été introduites dans le but de définir et de mettre en œuvre des politiques communes économiques, sociales et politiques. Toutes ces innovations ont contribué à créer le "peuple européen", au-delà des peuples nationales individuelles, avec des traditions et des institutions culturelles communes. En fait, chaque peuple est à la fois une population définible en termes de structure démographique et des dynamiques naturel et migratoire. Par conséquent, on peut se demander si dans le temps toutes les populations nationales de l'UE ont tendance à converger vers une "population européenne" avec un profil démographique commune. Dans cet article, je propose une méthode statistique pour mesurer la convergence des populations européennes vers une population commune.