Provisional draft

Prepared for presentation at Population Association of America, Boston, 1-3 May 2014

Demographic Trends in Developing Countries: Convergence or Divergence Processes?

Isabella Corazziari (*), Giuseppe Gabrielli (♦), Anna Paterno (♥), Silvana Salvini (♣)

(*) Italian Institute of Statistics, Rome (♦) University of Naples Federico II
(♥) University of Bari (♣) University of Florence

1. Introduction and aim of the study

Recent years are characterized by both a rise in life expectancy in the developing countries (DCs) and a further fall in fertility. According to this framework, the trends of demographic behaviours would have changed toward common patterns together with many modifications of social aspects (Salvini, 2004). Moreover, various theories of social change shared the assumption that societies would converge toward a condition of similarity (Coughlin, 2001). In this context processes of economic and behavioural globalization have been also ascertained.

In particular, the convergence theory argued that, as countries achieve similar levels of economic development, they become more similar in terms of these and other aspects of social life. This theory refers to the hypothesized link between economic development and concomitant changes in social organization. Nevertheless, these processes may coexist with a large heterogeneity in the living conditions and huge economic gaps among countries.

Overall, in most of population studies, the concept of convergence is linked to the demographic transition theory (Chesnais, 1988). Observing the experience of developed countries, the essence of this theory expects that fertility and mortality rates covary over time in a predictable and uniform manner. In this approach, the transition from a high fertility and mortality situation to another one characterized by low vital rates may be seen as an example of convergence, that describes a world going towards a new "demographic equilibrium" (Wilson 2001).

The aim of our research is to analyse the trends of specific demographic parameters regarding mortality and fertility, jointly with some socio-economic characteristics (living condition, socio-sanitary situation) of more than 100 DCs, to assess if convergence patterns in demographic behaviours prevail or if marked differences persist. As the paths of mortality and fertility in fact differ deeply over space and time, we need a specific statistical multi-way analysis technique that consider the time series dimension.

Since 70s' a growing interest has been given to multiway data, classified according to more than two dimensions (the classic units x variables), and many methods have been developed (Coppi, Bolasco 1989, Coppi 1994). Few methods consider a specific statistical treatment of the third dimensions (be it time or space or another criteria of classification), usually considered symmetrically with respect to the other two. Also for the methods that treat asymmetrically the third dimension, the ordering feature of time is not considered, with very few exceptions (Corazziari 1999). The Dynamic Factor Analysis has been proposed and developed in '70s (Coppi, Zannella

1979) for multiway data of the type unit x variables x time, considering explicitly the third ordered dimension representing time. The method is based on the joint application of a factorial analysis and regression over time to centers of specific dimensions. The method allows a descriptive and explorative analysis of data.

We apply Dynamic Factor Analysis (DFA) and Cluster Analysis of trajectories in order to evaluate at macro-level the main demographic trends of DCs in the 1995-2010 period. Results let us reconsider the processes of convergence and enlighten the heterogeneity among clusters. In paragraph 2 we make an overview of the theoretical and empirical literature on convergence and demographic transition. In paragraph 3 we describe data and methods used to analyze the trend of DCs to the aim of detecting the process of convergence/divergence in act. Paragraph 4 includes the results of our research and finally paragraph 5 contains discussion of results taking into account the initial considerations on convergence and demographic transition.

2. Convergence and demographic transition: theoretical and empirical literature overview

2.1. Theoretical background

In the past two hundred years, many social scholars hypothesized that the differences among societies would decrease over time (Inkeles, 1999) and that demographic behaviors would converge in the future (Wilson, 2011). In the 1950s and 1960s, the hypothesis of convergence has been associated primarily with the theories of modernization, which have generally assumed that developing countries would follow a path of economic and social progress similar to that of developed ones. The term modernization refers to the set of processes of change on a large scale that involves a particular society, profoundly transforming its structures and patterns of social organization. This concept refers more specifically to the trend of a society affected by these processes to acquire the economic, political, social and cultural characteristics typical of modernity, which therefore reflect aspects such as individualism and rationalism. The modernization is also closely related to the concept of economic development while the social dimension of modernization manifests itself in phenomena related to demographic change such as urbanization and extensive migration processes.

Other large-scale social transformations are included in modernization, such as population passages from traditional societies characterized by high levels of mortality and fertility (a demographic situation called “Ancien Régime”) to modern demographic regimes where vital rates are low. This passage is defined “demographic transition”, essentially related to the sociological concept of convergence, implying also the transformation of woman status, the so-called “female empowerment”.

More recently, the study of “post-industrial” society and the debate over “post-modernist” aspects of contemporary society also reflect the idea that there is a tendency for broadly similar conditions or attributes to emerge among a range of otherwise distinct and dissimilar societies (Salvini, 2004). As well as population dynamics are concerned, Chesnais (1988; 1997) and Oeppen (1999) were among the first to deal with convergence explicitly, while Heuveline (1999) considered the consequences of convergence on a regional and global scale. Similarly, the United Nations (http://esa.un.org/wpp/index.htm) bases their projections on the assumption of convergence, anticipating a homogeneous world in which almost all the demographic variability has disappeared (Wilson 2001 - Pop Dev Res).

From a methodological point of view, demographers draw on theoretical, methodological, and empirical literature, developed mainly by economists, where convergence lies at the heart of modern growth theory (Barro and Sala-i-Martin, 1992; 2004) and they applied a variety of statistical methods to test for convergence within and between countries.

Two of the most common methods are referred to as beta-convergence and sigma-convergence. Applied to demographic behaviours beta-convergence occurs when countries that are laggards in the demographic transition show more movement toward convergence than those further along in that process while sigma-convergence describes the overall spread of the observed distribution and refers to a reduction of disparities among countries in time (Sala-i-Martin 1996; Neumayer 2004). Referred to fertility and/or mortality, sigma-convergence implies that if the repeated cross-sectional standard deviation increases faster relative to the mean of the observed phenomenon, countries are diverging, and if the variance declines faster relative to the mean, countries are converging.

A third method, using measures of inequality, estimates the spread of a distribution (Firebaugh, 2003). With reference to fertility, inequality refers to the relative differences across countries in national TFR estimates, basing on the idea that an absolute decline in fertility gaps is neither a necessary nor a sufficient condition for declining inequality in fertility (Dorius, 2008). Lastly, we can remember the so-called “convergence clubs” approach, derived from the economic Solow model. This approach refer to groups of countries that show common trends, even if they differ from more general patterns of convergence (Sala-i-Martin, 1996; Dorius, 2008; Solow, 1956). This approach lies on the idea of conditional convergence. The equilibrium that each “club” will reach depends on the initial position and/or on other specific factors. The extended version of the Solow model (Lehmijoki and Pääkkönen, 2006) assumes that convergence should arise in demographic homogenous samples of countries, and that economic growth should be sensitive to demographic growth.

2.2. Empirical findings

With reference to developing countries, Wilson (2001 - Pop Dev Res) observed the trends of life expectancy at birth and of total fertility rate. He highlighted that the second half of the twentieth century witnessed a steady increase in the share of the world’s people living under conditions of declining fertility and rising life expectancy, describing this process as “global demographic convergence”. More in detail, he defined convergence in fertility as substantial, even if, in 2000, a significant tail of high-fertility populations remained, particularly in sub-Saharan Africa and in the northern part of Indian subcontinent. Similarly, he described an even more remarkable convergence in mortality levels. Lastly, observing that social and demographic change had progressed more rapidly than economic development, he saw the demographic convergence as one aspect of increasing social similarity.

Other analyses have dealt with convergence issue both in fertility and in mortality. Casterline (2001) modeled the pace of fertility decline in less developed countries from 1950 to 2050 and found a significant level of intercountry and intraregional variation in the pace of fertility decline. Wilson and Pison (2004), observing the cumulative distribution of the world’s population by fertility level in 1950-2003, suggested that, despite significant change in the middle of the distribution, the overall range does not declined. Reher (2004; 2007) highlighted that in much of the developing world the existence of prior and more rapid decline of child mortality played a central role for fertility decline. Dorius (2008) argued that the observed variation in intercountry fertility decline for much of the last 50 years pointed to divergence, rather than convergence and that countries began to converge only around 1995. He showed that the fall of fertility rates around the world does not necessarily meant that fertility rates were converging, defining the delayed onset of the fertility transition for many DCs as the single biggest source of divergence in the Total Fertility Rate (TFR). He identified a possible explanation for convergence in health, wealth, and life expectancy with the consistent link between economic and social development, and considered fertility less consistently linked to development than other variables have.

A growing body of empirical research generally confirms the occurrence of rising and converging life expectancy levels. In fact, a number of studies demonstrated that most of the period from 1920 to 2000 was one of convergence for many countries (Becker, Philipson, and Soares 2005; Bourguignon and Morrisson 2002; Easterlin 2000; Goesling and Firebaugh 2004; Neumayer 2003, 2004; Pradhan, Sahn, and Younger 2003; Ram 2006). Notwithstanding, some researchers defined convergence as "modest" at any time (Moser, Shkolnikov, and Leon 2005). McMichael (2004) suggested a recurring transition process of health, observing that reversals have occurred in many countries in recent decades due in large part to declining male life expectancy in Eastern Europe and Russia and the spread of HIV-AIDS primarily in sub-Saharan Africa. In addition, Moser (2005) confirmed that a long period of global convergence in life expectancy at birth has been replaced since the late 1980s by divergence of some regions of the world (such as many sub-Saharan countries or some eastern European countries), despite the improving of global life expectancy at birth in the period 1950-2000. Moreover, he showed that the shift from global convergence to divergence has been driven by reverse in adult mortality and that divergence appeared relatively small and of limited duration compared with the earlier convergence. Other researchers showed, in a cross-country perspective, that convergence in demography might occur conditioned on comparable socio-economic and environmental characteristics (Mishra, 2011). Considering different variables, Angeli and Salvini (2009) carried out a descriptive analysis of population characteristics of countries at low and medium Human Development Index (HDI) levels, showing that some exceptions emerged in the convergence process in the mean values of parameters. The same authors highlighted both the strong link between social, economic and demographic development, and the need to examine the paths taken by different countries on the way of the globalization of behaviours.

Recently, Wilson (2011) viewed most demographic change over the past half century as falling along a "main sequence" of demographic transition and the large majority of the world's population as engaged in a process of convergence. The principal differences between the regions of the developing world are identified in when they enter this main sequence and how rapidly they move along it. Observing the similarity of the regional trends along the main sequence, this author suggested that the health and fertility transitions are tightly connected. In particular, he considered the fertility transition as a truly global process, with no evidence of significant reversals and only a few countries still to embark upon it. In contrast, he defined the health transition as a slower transformation characterized by a "disturbing" evidence of its fragility, with stagnation and reversals affecting hundreds of millions of people. Moreover, he showed that, when considering the progress of the health transition, the world is not a single demographic system, but it is divided by deep faults into a number of blocs, each with its own distinctive trajectory of life expectancy.

Our research fits into this theoretical and empirical framework. We will consider the recent trends of socio-demographic parameters of DCs to understand if a process of convergence is in act and if clusters of countries in terms of time trajectories of demographic behaviors may be outlined and interpreted.

3. Data and methods

The analyses are carried out on 105 countries with a population of at least 1 million in 2010, defined by UN "less" or "least" developed countries. Data in analysis come from major international sources (World Health Organization, World Bank, United Nations), and they refer to the main vital events and socio-economic indicators for the DCs. Unfortunately it was not possible to consider the years prior to 1995 for the presence of too many missing data in the considered variables.

We analyse separately fertility and mortality processes, measured through their main indicators, but on the same set of countries, so three of the initial ones (Lesotho, Sierra Leone and Panama) have been excluded from the analysis as fertility and mortality indicators were not fully available. Both fertility and mortality have been associated in the analysis with the most commonly correlated socio-economic variables (i.e. for fertility, contraceptive prevalence, gender parity index in school enrolment, etc., and, for mortality, immunization coverage among infants, access to improved sanitation, etc.) that may contribute to tracing different "patterns" in observed DCs.

To study the process of convergence/divergence of the temporal dynamics of fertility and mortality, we use a method for multi-way data based on the joint application of a factorial analysis and regression over time called Dynamic Factor Analysis (DFA). The method has been developed in 1970s by Coppi and Zannella (Coppi et al 1979) and released in 1990s by Corazziari (Corazziari 1999).

DFA considers quantitative array of data classified according to the following three criteria (or modes, see Tucker 1966): statistical unit, quantitative variable and time of data collecting. This kind of data may be represented in a cubic matrix X (Law and others, 1984) whose generic element is

$$\boldsymbol{X(I, J, T)} = \{x_{ijt}\},\ i=1,\ldots I,\ j=1\ldots J,\ t=1\ldots T$$

where *i* is the unit index, *j* the variable index and *t* the time index, and the same units and variables are observed in each time (or occasion). Broadly speaking, this kind of methodology manages to combine, from a descriptive point of view, the Principal Component Analysis of a compromise matrix over time, and the analysis of the time dynamic of the array by linear regression models of polynomials in *t* of *k* order.

In the cubic array $\boldsymbol{X(I, J, T)}$ three sources of variation can be considered and modelled, each of which depending on the two modes of the arrays that can be considered (units x variable; variables x time; units x times). Weights for each dimension of the array can also be considered (weights for units, variables, times). The first source of variation can be attributed to the joint interaction of variables and units, a sort of structural variability or *static*, that is the undertone of the overall variability subject to time changes due to time interaction with variables and with units.

The second and the third sources of variation refers to time and its relation with units and variables. In particular, the dynamic of variables over time is represented by the variability of the mean of each variable over time $(x_{.jt})$; the dynamic of units is represented by the time changes of a barycenter of each unit over the set of variables (an average of variables for each unit). As the focus of the present work is on the indicators of fertility and mortality, more relevance is given to variables and their dynamics[1], so the units' one is considered as differential: given the mean time changes of each variable, each unit will be observed in its net time variation, that is if it strengthens the change of the variables or it moves in other directions over time, weakening the overall dynamics.

The goal of the methodology is to linearly decompose the overall variability characterising the observed data, described by the covariance matrix of $X(I, J, T)$[2] in the three sources of variability described above, called static (a sort of mean over time), dynamic of centres $(x_{.jt})$ and units' differential dynamic (the net dynamic of single units, when the centres trends have been subtracted). It has been shown (Coppi 1979; Corazziari 1999) that the overall covariance matrix may be decomposed into the sum of three covariance matrices each of them describing one of the above sources of variation:

$$\boldsymbol{S} = {}^{*}\boldsymbol{S}_i + {}^{*}\boldsymbol{S}_t + \boldsymbol{S}_{it}$$

where S is the overall covariance matrix of the array $\boldsymbol{X(I, J, T)}$, $^{*}\boldsymbol{S}_i$ is the covariance matrix of the centres $x_{ij.}$ representing the mean structure over time, $^{*}\boldsymbol{S}_t$ is the covariance matrix of $x_{.jt}$, and $\boldsymbol{S}_{it}$ is

[1] The DFA provides the possibility to consider both the dual and the tridual extension of each of the four models it considers, according to which dimension is considered more relevant and strategic in the analysis. The dual version focuses on the units dimension, and the tridual on time dimension.

[2] In Corazziari (1999), it has been shown that the overall covariance matrix of X(I,J,T) is the covariance matrix of the two-way matrix X(IT,J), obtained collapsing the single matrices $X(I,J)_t$ in each time, over time.

the covariance matrix representing the differential dynamics of units, after subtracting the mean variables dynamic and the static source of variation.

The DFA consists in 4 models each of which approaching the three sources of variation with a specific strategy.

As regards time evolution of the centres $x_{.jt}$ all of the three AFD models consider a linear regression model for each variable j, where the independent variable is time. The parameters are obtained by ordinary least squares. The assumptions about residuals $e_{.jt}$ are the classic ones: $cov[e_{.jt},e_{.j't'}]=w_j$, if j=j' e t=t', and 0 otherwise.

The variability of the centres $x_{ij.}$ is analized by factorial analysis of specific covariance matrices in each of the three AFD models.

In the first model, factorial analysis is applied to the covariance matrix $\mathbf{S_t={}^*S_i + S_{it}}$, and it can be easily shown that $\mathbf{S_t}$ is also obtained as the sum of the covariance matrix of variables in each occasion, divided the total number of times. So by projecting the matrices $\mathbf{X_t}$ centred in each time we obtain the factorial representation of each unit in each time, that is their dynamic, net to the overall mean time dynamic. The representation of the centres $x_{ij.}$ is obtained by projecting their matrix $\mathbf{{}^*X_i}$ centred, on the factorial plane, due to the decomposition of $\mathbf{S_t={}^*S_i + S_{it}}$, as $\mathbf{{}^*S_i}$ is the covariance matrix of $\mathbf{{}^*X_i}$. The other models considers different matrix for the analysis of the static source of variation, and for the analysis of the differential dynamic of units (Corazziari 1997;1999).

As regards differential time evolution of the units, in the first model it is described by comparing the projection of each unit in each time, with the projection of the corresponding centre $x_{ij.}$, according to the decomposition $\mathbf{S_t={}^*S_i + S_{it}}$. Both the second and the third AFD models describe the differential time evolution of the units, starting from a time regression model for each unit, whose parameters are calculated by ordinary least squares:

$$\mathbf{x_{ijt}=a_{ij}+b_{ij}\cdot t+e_{ijt},\ j=1,...,J \text{ and } i=1,...,I}$$

the assumptions are: $\mathbf{cov[e_{ijt},e_{ij't'}]=w_j}$ **if j=j' and t=t'**; 0 otherwise.

Differential time evolution of each unit can be measured considering the differences of the two regression parameters $\mathbf{b_i}$ and $\mathbf{b_{ij}}$, **j=1,...,J**.

Indexes of the goodness of fit of each source of variation in each models are also provided. They are calculated as the ratio between the trace of the modelled covariance matrix of the specific source of variation, and the corresponding observed trace, for each of the covariance matrices described above.

In the following two applications (one for mortality and the other for fertility), the first model of the DFA has been applied. The units are the countries. For the analysis of mortality the variables are 8 yearly indicators (death probability at age 15-60 separately for males and females, child mortality, HIV-AIDS prevalence, HDI, immunization for DPT3 - diphtheria, pertussis and tetanus -, access to improved sanitation and a macro area variable (that is the geographical region of belonging)[3]. For the analysis of fertility, 6 yearly indicators have been considered (adolescent fertility rate, TFR, contraceptive prevalence, HDI, gender parity index in secondary school enrolment and macro area).

Times are the 4 years 1995, 2000, 2005 and 2010. Each application identifies a factor plane whose interpretation is based on the correlation coefficients between the variables and the axes of the factor plane. The trajectories of the projected countries over the plane are then analysed by a cluster analysis of trajectories (Carlier 1986) obtaining clusters of countries homogeneous as

[3] The macro-area is conceived as a proxy variable summarising other characteristics of each country not expressed by the chosen indicators, but important to limit clustering of countries too far geographically. It is conceived as a variable considering both homogeneity of countries belonging to the same macro-area, than, at a lesser extent, proximities of different macro-areas, that is macro-area with value 1 is more close to macro-area with value 2 than other macro-areas for example. Also a disjunctive set of variables each referring to a single macro-area has been considered in alternative, providing similar but more scattered results, so the macro-area unique variable has been preferred in the analysis.

regards the level and dynamics of the considered variables, more easily interpretable in their positions over the plane.

The main feature of the cluster analysis of trajectories is the type of distance between units. When studying trajectories two type of distance between couple of units can be considered: a mean of the comparison (differences) between the two units in each occasion (*mean instantaneous distance*) and a mean of the comparison of the variations between adjacent occasions of each unit (*mean unfolding distance*). A mean of the two considered distances has been proposed, with suitable weights giving more relevance to one of the two distances.

Given the above distances between units, the standard methods of cluster analysis is applied, defined on the basis of the defined distance between clusters.

In the present work, for both the mortality and fertility analysis, the mean of the two distances proposed by Carlier (1986) has been considered, and the Ward method has been chosen among the hierarchical cluster analysis methods, confirmed by a final k-means cluster analysis based on the barycentre of the clusters of the Ward better partition.

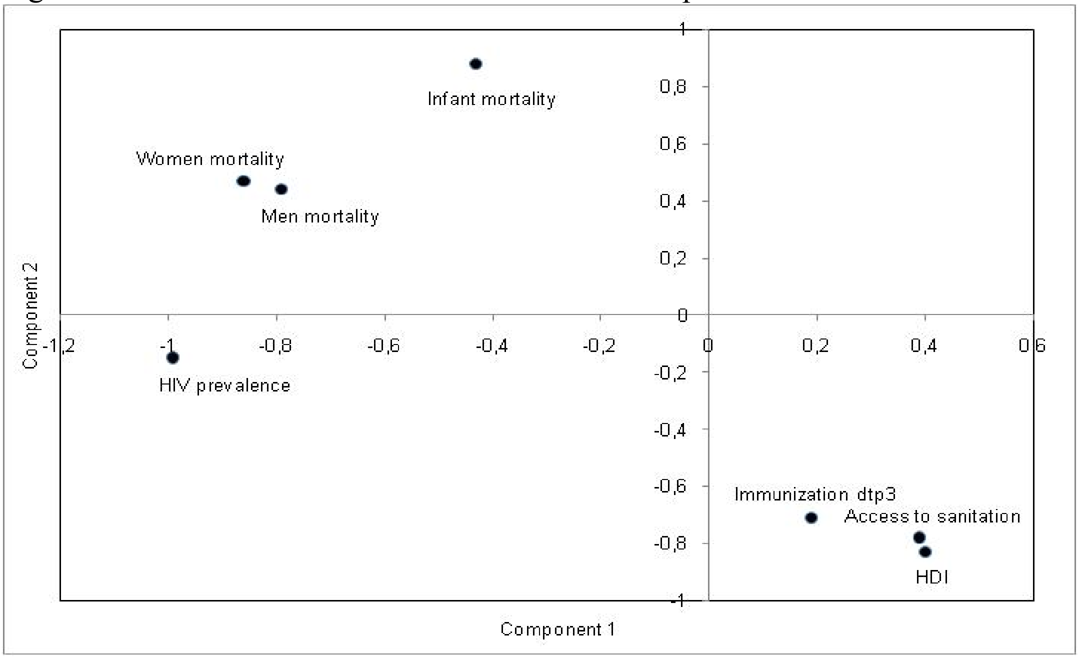
The trajectories of the units over the factorial plane have been clustered. If the clusters' trajectory tends towards the centre of the plane, characterising the overall average dynamic of the system of data, the clusters tend to buy homogeneity (witnessing the presence of a convergence process); if they move away from the centres, the heterogeneity increases.

4. Results

4.1. Mortality and health

DFA results show first of all that the first two components of the factor analysis explain a great part of variability (I_t=94.8%) and the best represented times are the second and the third ones (respectively 2000 and 2005, with percentages equal to 95.5 and 95.1). With regard to the correlation between variables and factors (fig. 1 and table 1) we note a strong negative correlation of the first component with both the probabilities of death in adult age (separately for men and women), and the prevalence of HIV-AIDS, while the correlation with the other variables is weaker and positive (except with child mortality). The second component is strongly and positively correlated with under-five mortality. The correlation is positive also with adult mortality (separately for men and women), but the values are low. Strong negative correlation of the second component is shown with HDI, the access to improved sanitation and immunization for DPT3. In synthesis, the first component assumes the meaning of mortality and morbidity (increasing values means lowering mortality and morbidity rates), while the second component represents sanitary conditions and overall health status of observed countries (increasing values of the components means a worsening of such conditions). Following this interpretation of our results, the positive correlation of the second component with child mortality can be understood if we consider the latter as an indicator of the health and socio-demographic development of populations.

Figure 1 - Correlation of variables with factorial components



Source: Our elaboration on data of World Health Organization, World Bank, United Nations

Table 1 - Correlation matrix between variables and the first two components.

Variables	Component 1	Component 2
Macro area	0.25	-0.34
Prob. of death 15-60 (males) - PDm	-0.79	0.44
Prob. of death 15-60 (females) - PDf	-0.86	0.47
Child mortality (M05)	-0.43	0.88
HIV-AIDS prevalence	-0.99	-0.15
Immunization for diphtheria, pertussis and tetanus (DPT3)	0.19	-0.71
Access to improved sanitation (AIS)	0.39	-0.78
Human Development Index (HDI)	0.40	-0.83

Source: Our elaboration on data of World Health Organization, World Bank, United Nations

The dynamic of centers ($x_{.jt}$) shows the trajectories of countries over the factorial plane; it is described through time regression of a suitable order as indicated in table 2. The overall index of fitness for this type of variability is good (0.971).

Only for variable HIV-AIDS prevalence, and the two indicators of adult mortality a polynomial in *t* (time) of second order has been required, for all the other variables (also for variables in the analysis of fertility) a simple linear regression model in *t* has been fitted. Average (over units) indicators of mortality decrease over the considered period, especially child mortality, while the HDI, the immunization DPT3 indicator, the access to improved sanitation increase. Also the HIV-AIDS mean prevalence decreases, mainly since 2000 (second occasion).

On the basis of the dynamic of trajectories, K-means results of cluster analysis have lead us to choose a partition formed by 7 clusters (both if using the hierarchical Ward methods, than the k-means confirmative one, with very similar aggregations of countries and very few exceptions). The sequence of the clusters expresses the ranking according to the mean value of adult mortality levels (from the highest to the lowest) and are here shown (see table 3). A geographic characterization is quite evident for 5 of them, while 2 (the first and the second one) include countries located in different macro-areas.

Table 2 - Time regression analysis of centres of units (overall index of regression fitness *It=0.971)

Variable	R-square	Least Square Estimates					
		Constant	(std. error)	Slope coefficient	(std. error)	Second order coefficient	(std. error)
PDm	0.992	1.068	(3.2E-2)	-1.5E-2	(5.8E-3)	1.8E-2	2.9E-2
PDf	0.979	1.037	(5.2E-2)	-2.0E-2	(9.5E-3)	4.5E-2	4.8E-2
M05	0.980	1.331	(3.7E-2)	-0.133	(1.3E-2)		
Hiv-aids	0.885	0.637	(0.136)	-6.6E-2	(2.4E-2)	0.344	(0.124)
DTP3	0.972	0.874	(1.7E-2)	5.0E-2	(6.1E-3)		
AIS	0.997	0.886	(5.2E-3)	4.6E-2	(1.9E-3)		
HDI	0.997	0.889	(4.9E-3)	4.4E-2	(1.8E-3)		

Source: Our elaboration on data of World Health Organization, World Bank, United Nations

Cluster 7 is the largest one, comprising 53 countries belonging to various macro-areas. This cluster is characterized mainly by the presence of low levels of adult mortality and child mortality, as well as for a scarce presence of HIV-AIDS. It appears also the best cluster with reference to immunization coverage, access to improved sanitation and Human Development Index. Therefore, it could be considered as including the "best" countries in terms of survival, health and socio-economic conditions. It may be considered also as a reference group when convergence dynamic is questioned. Cluster 6 includes 15 countries, also in this case belonging to various macro-areas. Its main features are levels of adult and mainly child mortality higher than cluster 7, although lower than the other groups of countries. Cluster 5 contains 20 Sub-Saharan countries, plus Afghanistan that shows an evident delay in the health transition. This cluster presents the lowest levels of immunization and of access to sanitation. It is characterized, above all, by high levels of child mortality (the high correlation with second component reported in table 1 shows this evidence) and one of the lowest Human Development Index.

Table 3 - List of analyzed countries by cluster of mortality (k-means method).

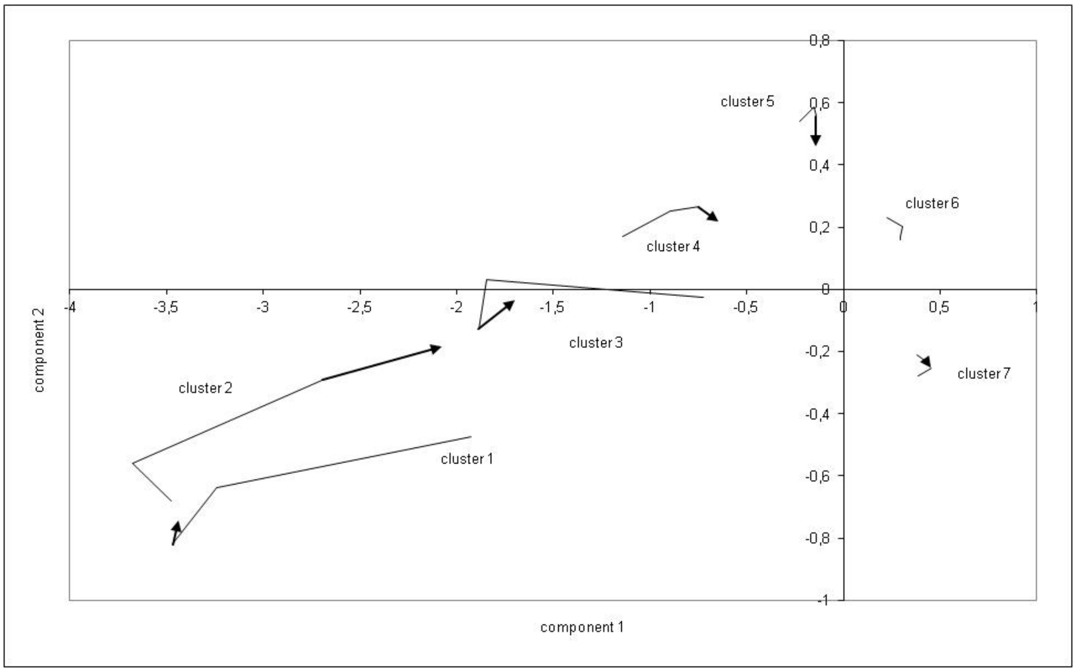
Cluster	Countries
1	Botswana, Swaziland.
2	Zimbabwe.
3	Malawi, Mozambique, Namibia, South Africa, Zambia.
4	Central African Republic (CAR), Cuba, Kenya, Tanzania, Uganda.
5	Afghanistan, Angola, Benin, Burkina Faso, Burundi, Cameroon, Chad, Congo, Congo Dem. Rep, Ethiopia, Gabon, Guinea, Guinea-Bissau, Haiti, Mali, Niger, Nigeria, Rwanda, Togo.
6	Bangladesh, Bolivia, Cambodia, Gambia, Ghana, India, Indonesia, Lao PDR, Madagascar, Mauritania, Mongolia, Myanmar, Nepal, Pakistan, Papua New Guinea (PNG), Senegal, Sudan, Tajikistan, Yemen.
7	Algeria, Argentina, Armenia, Azerbaijan, Belarus, Brazil, Chile, China, Colombia, Costa Rica, Côte d'Ivoire, Dominican Republic, Ecuador, Egypt, El Salvador, Georgia, Guatemala, Honduras, Iran, Iraq, Jamaica, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, Lebanon, Libya, Malaysia, Mauritius, Mexico, Moldova, Morocco, Nicaragua, Oman, Paraguay, Peru, Philippines, Qatar, Russian Federation, Saudi Arabia, Sri Lanka, Suriname, Syria, Thailand, Trinidad and Tobago, Tunisia, Turkey, Ukraine, Uruguay, Uzbekistan, Venezuela, Vietnam.

Source: Our elaboration on data of World Health Organization, World Bank, United Nations

Cluster 4 is formed by only 4 Sub-Saharan countries plus Cuba. They are featured by medium-high mortality levels, and, with reference to the other variables, by an intermediate position. Cluster 3 comprises 5 Sub-Saharan countries characterized by the second highest child mortality rates despite the presence of medium levels of immunization and access to sanitation. The only country present in cluster 2 is Zimbabwe; it is characterized by a very high adult mortality and prevalence of HIV-AIDS. The same disease is widespread in Cluster 1, that is formed by only 2 countries (Botswana and Swaziland). Although the last 3 countries share many aspects of their mortality and morbidity conditions, they are divided in different clusters due to the different dynamic showed by this aspects over the time, as can be noted observing fig. 2.

According to the DFA method, differential dynamic of median centers of clusters is represented in figure 2. Generally speaking, the representation of the trajectories of the clusters on the factor plane shows a slight trend toward the average situation; in other words, we observe weak differential dynamics towards the center of the axes. Cluster n.1, 2 and 3 are located in different quarters of the factorial plan, according to the different values of the active variables. However, looking at the temporal trends of each cluster, we observe that they tend to slightly converge toward the center of the axes of the factorial plan. In other terms, the dynamic of countries included in such clusters (that represent almost the 90% of the global scenario included in the analysis) show a slow reduction of their differentials toward an uniformity to the centers of considered variables.

Figure 2 - Differential dynamics of median centers of clusters – Health and mortality



Note: cl_i means cluster i, t_i means time i.
Source: Our elaboration on data of World Health Organization, World Bank, United Nations

As to cluster 6 and cluster 4, the pattern of convergence toward the centre of the axes is much more evident. They represent those countries in which HIV-AIDS prevalence significantly decreased in the considered period: respectively from 25.0 to 15.2% in cluster 7 and from 9.3 to 5.8% in cluster 6.

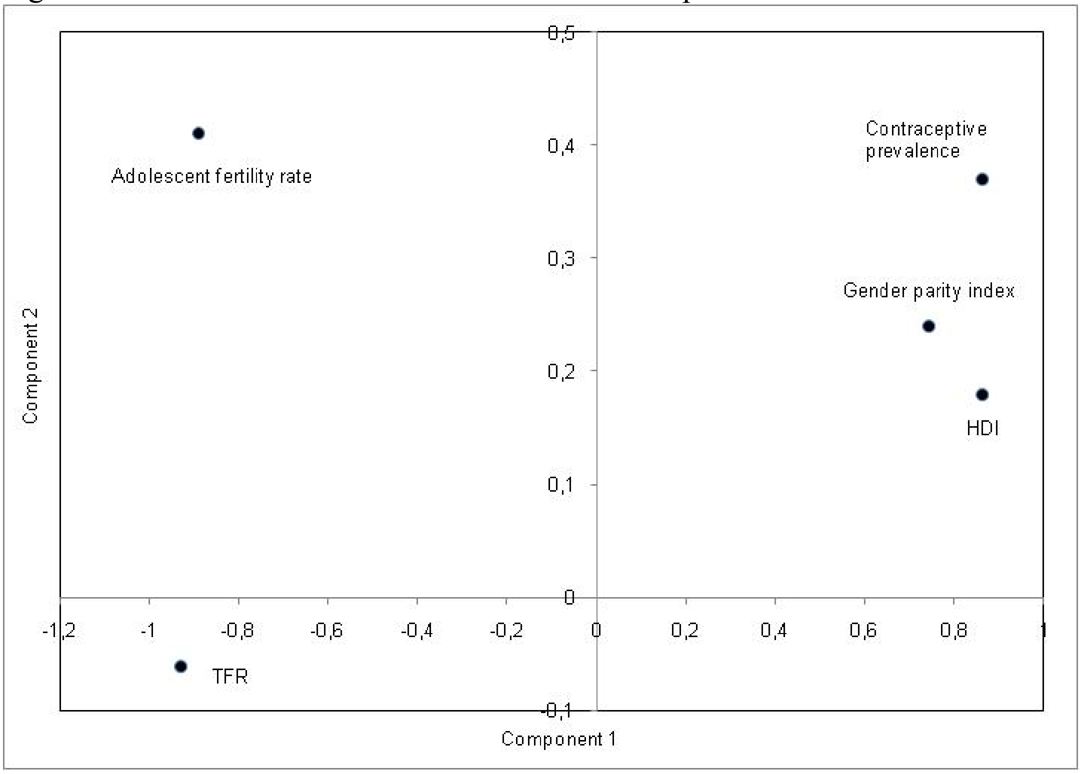
Clusters 5 and 7 seem to outline a worsening of their conditions, showing a differential dynamic that moves toward left in the plane that is toward highest levels of the aforementioned variables. This trend seems to show an opposite tendency with respect to the decreasing mean of the general mortality and of the HIV-AIDS levels, as indicated by the regression analysis of such indicators centers over units (table 2). However, both clusters shows a slight but significant “reversed” path dynamic in the period 2005-2010, oriented toward the center of the axes.

4.2. Fertility

Our results of fertility analysis can be synthesized as follows. The first two components of DFA explain 85.4% of the variability of the phenomenon, a little less than in the mortality analysis where, nevertheless, variables are slightly more numerous. For fertility, the times better represented are the third and the fourth (respectively 2005 and 2010, with percentages of explained variability equal to 86.5 and 85.6).

The correlation of the two components with the active variables (see table 4), indicates that the first component represents with positive values better situations in terms of human development, gender parity index of education, and contraceptive prevalence, while with negative values higher total and adolescent fertility rates (figure 3).

Figure 3 - Correlation of variables with factorial components



Source: Our elaboration on data of World Health Organization, World Bank, United Nations

Looking at the factorial plan, that reports graphically the correlations between factor components and variables exposed in table 4, we can note the separation described above between adolescent fertility and total fertility rate from one side, and GPI, HDI and contraceptive prevalence from the other. In synthesis, the variables measuring development are opposite to fertility behavior. This factorial plan is useful to understand the trajectories of clusters in the period 1995-2010.

Table 4 – Correlation matrix between variables and the first two components.

Variables	Component 1	Component 2
Macro area	0.28	0.75
Adolescent fertility rate	-0.89	0.41
Total Fertility Rate (TFR)	-0.93	-0.06
Human Development Index (HDI)	0.86	0.18
Gender parity index in school enrolment	0.74	0.24
Contraceptive prevalence	0.86	0.37

Source: Our elaboration on data of World Health Organization, World Bank, United Nations

To better describe the trajectories of countries over the factorial plane, a look to the overall dynamic of the data is necessary. The following table reports the time regression parameters of the considered indicators means.

Table 5 - Time regression analysis of centers of units (overall index of regression fitness *It=0.989)

Variable	R-square	Least Square Estimates			
		Constant	(std. error)	Slope coefficient	(std. error)
Adolescent fertility rate	0.989	1.228	1.850E-2	-9.134E-2	6.757E-3
Total Fertility Rate (TFR)	0.984	1.229	2.292E-2	-9.171E-2	8.369E-3
Human Development Index (HDI)	0.997	0.889	4.918E-2	4.430E-2	1.796E-3
Gender parity index in school enrolment	0.999	0.939	1.178E-3	2.437E-2	4.301E-4
Contraceptive prevalence	0.991	0.801	1.481E-2	7.975E-2	5.408E-3

Source: Our elaboration on data of World Health Organization, World Bank, United Nations

All the indicators of fertility are decreasing over time, while indicators describing development are increasing on average in the considered period. According to such dynamics, the positions and trajectories of the aggregated clusters are to be described (table 6).

The cluster analysis of the countries trajectories projected on the plane have lead us, also for fertility, to choose a partition formed by 7 clusters (again both if using the hierarchical Ward method and the k-means confirmative one, with very similar aggregations of countries and very few exceptions). The K-means results are shown here, where the number of the clusters expresses the ranking according to the mean value of TFR (from the highest to the lowest), generally describing the demographic transition stage.

The chosen partition shows well separated clusters of comparable magnitude (in particular for the clusters 2, 5, 6 and 7, respectively compounded by 17, 19, 20 and 18 countries). The geographic characterization is quite marked for the cluster 2 and 5 that include Sub-Saharan countries, with the only exception of Afghanistan and Yemen. These countries, as sub-Saharan Africa ones, are characterized by a strong delay in the demographic transition, that is high levels of fertility and adolescent fertility and a low value of HDI. Note that adolescent fertility, strongly related to women empowerment and female human capital, may be considered a proxy of (under) development of women status and, consequently, of the degree of growth of the country. The cluster 3 includes countries belonging to Southern Asia and sub-Saharan Africa, facing a medium stage of the path toward modernization. The cluster 4 is geographically more heterogeneous, including three Asian countries (India, Bangladesh and Nepal) but also some African countries, and Guatemala and Haiti, in Central American and Caribbean regions; it is characterized by high level of adolescent fertility. The cluster 5 is compounded by Latin-American countries, where fertility is lower, contraception higher, such as the development indicators HDI and GPI. Cluster 6 and 7 contain countries belonging to different continents that, nevertheless, present common demographic characteristics. Cluster 6include several Asian countries that in the recent past made part of Soviet Union, and that actually are living the process of economic and demographic transition. In cluster 7, China and Iran are an example of reached fertility transition, with value of TFR that nowadays are below replacement level. In this cluster the region mostly represented is the southeastern shore of Mediterranean, where demographic characteristics approach those of European countries (i.e. Lebanon, Tunisia, Turkey), even if the stage of development and women’s status described by HDI and GPI are quite different.

Table 6 - List of analyzed countries by cluster of fertility (k-means method).

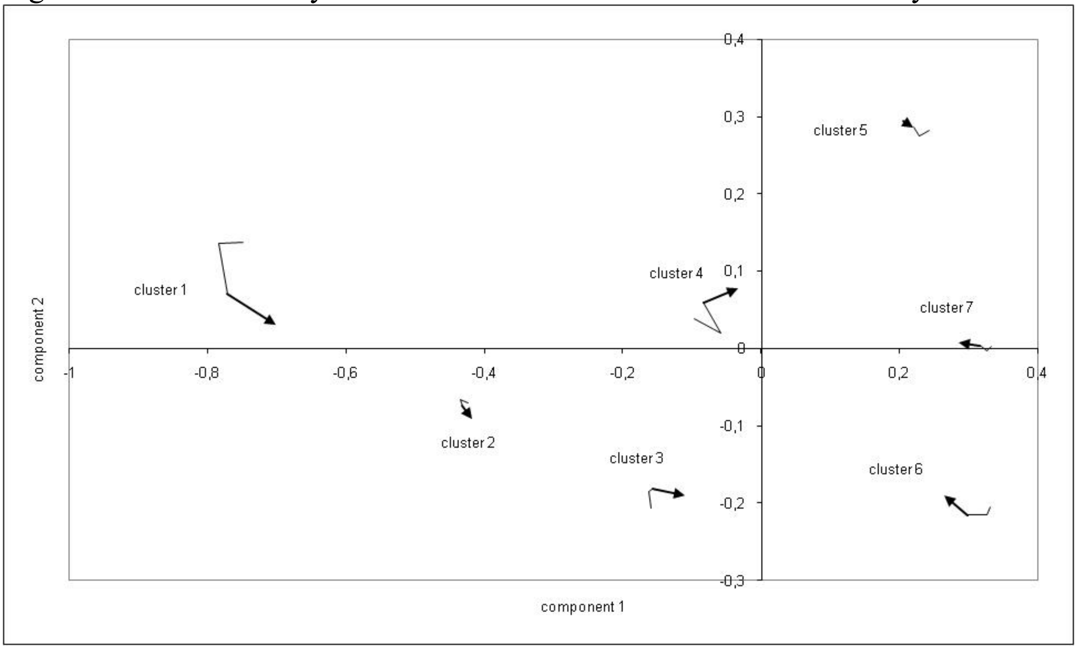
Cluster	Countries
1	Afghanistan, Angola, Chad, Congo Dem Rep, Guinea, Mali, Niger, Uganda.
2	Benin, Burkina Faso, Cameroon, Central African Republic (CAR), Côte d'Ivoire, Guinea-Bissau, Madagascar, Malawi, Mozambique, Nigeria, Senegal, Tanzania, Yemen, Zambia, Congo, Ethiopia, Gambia.
3	Burundi, Cambodia, Ghana, Lao PDR, Mauritania, Pakistan, Rwanda, Sudan, Togo.
4	Bangladesh, Gabon, Guatemala, Haiti, India, Iraq, Kenya, Namibia, Nepal, Papua New Guinea (PNG), Swaziland, Zimbabwe.
5	Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Honduras, Jamaica, Mexico, Nicaragua, Paraguay, Peru, Uruguay, Venezuela.
6	Algeria, Armenia, Azerbaijan, Belarus, Egypt, Georgia, Kazakhstan, Kyrgyzstan, Libya, Malaysia, Moldova, Morocco, Myanmar, Russian Federation, Tunisia, Ukraine, Uzbekistan, Tajikistan, Saudi Arabia.
7	Botswana, China, Indonesia, Iran, Jordan, Kuwait, Lebanon, Mauritius, Mongolia, Oman, Philippines, Qatar, South Africa, Sri Lanka, Suriname, Syria, Thailand, Trinidad and Tobago, Turkey, Vietnam.

Source: Our elaboration on data of World Health Organization, World Bank, United Nations

In synthesis, cluster 5, 6 and 7 present a better situation in terms of index of development and use of contraception; in opposite, clusters 1 and 2 are represented in the left part of the diagram, manifesting a worse situation in terms of development status.

In figure 4 we have reported the trajectories of median centers of the clusters for fertility analysis. We observe the temporal dynamic of the clusters with respect to the center of the axes, representing the reference of the overall dynamic on average. The clusters 4, 6 and 7 tend to converge toward the mean situation of the whole group of countries in the period we examine, represented by the barycenter of the axes. This means that the differential dynamic of countries included in these cluster tends to uniform to the dynamic of the centers of variables that determine the factorial plan. Cluster 1 starts from the worst position and goes towards the baricentre, Cluster 2 seems more stable over time and cluster 3, more close to the baricentre, does not show high variability over time as well. Overall, the picture described in figures 3 and 4 put our countries on a plan connecting variables and time trajectories of the clusters so to permit us to understand the paths toward demographic transition, development stages and convergence or divergence.

Figure 4 - Differential dynamics of median centers of clusters – Fertility.



Note: cl_i means cluster i, t_i means time i.
Source: Our elaboration on data of World Health Organization, World Bank, United Nations

5. Provisional and synthetic discussion

The aim of the paper was to analyze the trends of specific demographic parameters regarding fertility and mortality, jointly with some socio-economic characteristics (living condition, socio-sanitary situation) of more than 100 DCs, to assess if convergence in their demographic behavior prevails or if marked differences remain. The analysis of mortality seems to confirm that a large part of developing countries are converging on a uniform model of health and mortality, leaving backward in particular the countries characterized by high values of HIV-AIDS prevalence and – consequently – high mortality. Almost all of these countries are localized in Sub-Saharan Africa. The analysis of fertility indicators lead us to different considerations: countries converge to different models characterized to the belonging to specific territorial contexts. The deepening of results will permit us major comments.

Main References (to be updated)

Angeli A., Salvini S. (2009). Demographic Heterogeneity of Developing Countries between Reality and Convergence Myth, *XLII Scientific Meeting of Italian Association for Population Studies*, Milano.

Carlier A. (1986). *Factor Analysis of Evolution and Cluster Methods on Trajectories*, Compstat 1986.

Coppi R., Zannella F. (1979). L'analisi fattoriale di una serie temporale multipla relativa allo stesso insieme di unità statistiche, *Atti della XXIX Riunione della SIS*.

Corazziari I. (1997). Dynamic Factor Analysis, Classification and data Analysis, Pescara 3-4 luglio 1997.

Corazziari I. (1999). *Analisi di Dati Longitudinali Multivariati*, Tesi di Dottorato XI ciclo, Napoli aprile 1999.

Dorius S.F. (2008). Global Demographic Convergence? A Reconsideration of changing Inter-country Inequality in Fertility, *Population and Development Review*, 34(3), 519-537.

Lehmijoki U., Pääkkönen J. (2006). Demographic Clubs: Convergence within, divergence between, *Helsinki Center of Economic Research, Discussion Papers*, 136.

Sala-i-Martin X. X. (1996). The classical approach to convergence analysis, *The Economic Journal*, 106, 1019-1036.

Salvini S. (2004). Demographic Convergence and Economic Divergence? Revisiting Demographic Transition in Developing Countries, *Proceedings of the XLII Scientific Meeting of the Italian Statistical Society* (SIS), Bari

Solow R. M. (1956). A Contribution to the Theory of economic Growth, *Quarterly Journal of economics*, 70, 65-94.

Wilson C. (2001). On the scale of global demographic convergence 1950-2000, *Population and Development Review*, 27(1), 155-171.

Wilson C. (2011). Understanding global demographic convergence since 1950, *Population and Development Review*, 37(2), 375-388.