TOWARDS AN ESTIMATE OF CHILD WELFARE

CER – Centro Europa Ricerche

October 2004

CONTENTS

ABSTRACT

The CER seeks in this work of empirical analysis to develop an indicator of child welfare by means of the well-known statistical methodology of principal components analysis. Following in the path of the work initiated by scholars such as Jacques Van der Gaag and Erika Dunkelberg, the analysis regards three essential aspects: the key components accounting for the variability of child welfare in the different countries, the comparative weight of the policy measures susceptible of implementation in the various major areas determining welfare, and the classification of the different countries considered and their breakdown into groups with homogeneous characteristics. It is precisely the results of the final classification, differing only slightly from those obtained by works of a more deliberately theoretical nature, that the authors see as providing confirmation of the suitability of the methodology employed and therefore suggesting that the fullest advantage should be taken of the most precious indications obtained, namely those regarding the combination of policies best able to improve the situation.

1. INTRODUCTION

Ever since the 1970s there has been a growing divide between the dynamics of income and the evolution of "quality of life". While the concept of quality of life naturally lends itself to many different definitions,1 including for example personal freedom of choice, healthy environment, and the quality of interpersonal relations, one fact is certain, namely that it cannot be so narrow as to include only the element of income produced and distributed.

It is for this reason that in the 1980s a series of empirical studies, first in the United States and then in Europe, began to develop alternatives yardsticks of welfare, including the Index of Sustainable Economic Welfare (Cobb, Daly, 1989), the Index of Social Health (Miringoff, 1999), and the Genuine Progress Indicator (Redefining Progress, 2001).

GDP remains a highly attractive indicator due to its characteristics of apparent objectivity and universality deriving from the consolidated statistical methodology through which it is constructed. It is difficult to recognise the same characteristics in more complex yardsticks of welfare given the inevitable increase in subjectivity the closer the concept of quality of life approaches that of happiness. The challenge must, however, be accepted. Otherwise there is a risk of situations perceived by common sense as very different being placed at the same level.

The above considerations also hold, of course, for efforts to construct an index of child welfare, which can in turn be regarded as contributing to the definition of a concise general indicator of human development and welfare.

If we are to get some idea of children's living conditions, the economic picture must be supplemented with other indicators such as the infant mortality rate, life expectancy at birth, and the rate of school attendance or juvenile crime. These are just some examples serving to underscore the fact that welfare is a multiform notion at the conceptual level and not always directly measurable in empirical terms.

---

[1] According to the Nobel laureate Sen, for example, quality of life tends to coincide with the possibility for individuals to engage freely in activities consistent with the values in which they believe while respecting the rights of others,

The identification of indicators is thus a delicate stage in the process of constructing a welfare index. Failure to include any element with a bearing on the phenomenon means distorting the results of the study and increasing the comparative weight of the variables considered disproportionately with respect to those omitted. The opinion of experts who have already succeeded in identifying the relevant thematic areas is thus of crucial importance in the selection of indicators.
The essential point of reference in this case was Jacques Van der Gaag and Erica Dunkelberg's attempt to develop a comprehensive child welfare index,2 which offers a broad overview of the studies carried out in this connection and also contributes toward identification of key indicators needed to represent the phenomenon. In particular, the authors pinpoint the theoretical reasons for including in the child welfare index the three dimensions considered relevant for assessment of the degree of human development in general. These dimensions regard economic factors (lifestyle and prosperity), culture (level of education and proportion of the population with access to schooling), and the "health and longevity" of individuals.

[2] *Measuring Child Well-Being in the Mediterranean Countries – Toward a comprehensive child welfare index*, Jacques van der Gaar, Amsterdam Institute for International Development, and Erika Dunkelberg, Consultant, HDNED World Bank.

2. THE IDEAL CHARACTERISTICS OF INDICATORS AND THE INDICATORS ACTUALLY AVAILABLE

We have stated our theoretical frame of reference, which is indispensable in addressing the preliminary problem of identifying a series of indicators. These must then in turn possess technical characteristics such as significance, widespread availability and homogeneity. Significance refers to the indicator's ability to provide a quantitative measurement of a specific aspect of the phenomenon. Availability is connected with the possibility of obtaining data with respect to all the statistical units considered. Finally, the indicator's homogeneity depends on the degree to which the different countries involved adopt the same survey criteria.

With respect to our analysis in particular, the high degree of heterogeneity characterising the statistical units of reference, i.e. the Mediterranean countries (including both advanced countries like France and less developed countries like Albania and Iraq), makes the balanced selection of indicators more difficult for at least two reasons. Firstly, the less developed countries may in fact also suffer some disadvantages in terms of the availability of information. Secondly and more importantly, it is not always correct to attribute the same significance to the same indicator in the presence of strongly differentiated socio-economic structures. We shall discuss below how this problem, which remains ineradicable in many respects, was tackled.

3. ANALYTICAL INDICATORS AND SYNTHETIC INDICATORS

In an ideal methodological pathway, once the key indicators have been identified and their effective availability and comparability ascertained, it is a matter of choosing between a more analytical and a more synthetic representation of the phenomenon in question.

In this connection, an initial representation of child welfare is unquestionably provided by the simple combination of many different measures. Each variable naturally expresses an individual aspect of the phenomenon. But how important are the individual aspects? It is not easy to say. While the link between the individual indicators and the phenomenon under examination is clearly defined, it is difficult to pinpoint objectively out of the different combinations the one associated with the greatest degree of welfare. There can be no doubt that access to drinkable water is a requisite for welfare, as are a low juvenile crime rate and a low percentage of adolescents undergoing abortions. It is, however, much more difficult to understand whether a child’s living conditions are better in a country where access to water is guaranteed but the economy is stagnating or in one where GDP displays relatively dynamic growth but water resources are not available throughout national territory. The attempt to construct an indicator taking into account some trade-offs between the variables considered certainly takes us away from an objective representation of the phenomenon but may offer some gains as regards indications for economic and social policy. It is for this reason that it may make sense to try and construct a synthetic indicator alongside the analytical ones.

In constructing a synthetic index it is possible to choose from a range of differentiated techniques with greater or lesser degrees of complexity.

First of all, it is necessary to choose between a weighted and a non-weighted index.

A non-weighted index involves calculating the simple mean of the indicators selected. The sole requisite is the existence of an unambiguous and clearly defined link between each numerical indicator and the phenomenon to be accounted for. For example, it must be clear that a higher level of per capita income is equivalent to a comparatively high degree of welfare, just as there is a positive correlation between child welfare and a lower number of underweight births. If these links hold for all the indicators, it will be sufficient to relate all the

information to the same unit of measurement, thus standardising the variables and translating them into pure numbers. Once this operation is completed, the arithmetic mean will represent a global value making it possible to rank each country in a general classification. The index thus obtained will represent child welfare as long as there are no serious omissions and that the set of variables selected exhausts all the information needed to explain the phenomenon.
When a non-weighted synthesis is produced, the classifications constructed on the basis of individual variables are essentially added or converted into means. The simplicity of the method in both conceptual and practical terms constitutes an advantage. In the absence of a weighting criterion, however, no relations are established between the variables and there is thus no examination of the trade-offs that can prove so interesting when framing policies.
The latter observation suggests the need to develop a new aggregative method that combines the individual classifications while weighting the variables examined.
There are the two primary methodological paths followed in the definition of weights. In the first, calibration is based on the subjective choices of experts or theoretical interpretations of the phenomenon (a economically oriented vision will attribute considerable weight to the economic variables, regarding the others as “superstructure” or an effect of economic development). The degree of subjectivity can be reduced with the aid of statistical analyses making it possible to distinguish between key variables and “superstructure” variables on the basis of their capacity to account for the phenomenon. The former can then be assigned the maximum weight in constructing the index.
Calibration takes place in the second case through a statistical methodology that directly identifies the weights to be assigned to each of the variables considered. It was decided to opt for this second approach, which can, as will be evident from the above considerations, be of assistance also in terms of enriching the first methodological approach.
To this end, use was made of factorial analysis or, to be more precise, analysis in terms of principal components, which constitutes its particular application to data of a quantitative nature. This methodology is normally used in the study of complex phenomena defined by the operation of multiple factors in different directions, and makes it possible to obtain three fundamental results, namely a preliminary, in-

depth study of the relations between the variables, the construction of a comprehensive indicator, and its use to establish a classification of the survey units (countries in this case).

The primary strength of factorial analyses is their ability to manage very complex databases in a unified fashion through synthesis of the information derived on the basis of the relations observed between the individual variables. Given the large number of factors involved in its definition, child welfare falls unquestionably within the group of phenomena susceptible of exploration by these methods.

While referring readers to the methodological appendix for a more complete and specific description of the technique in question, we shall seek to outline its primary properties. On the sole assumption that child welfare in the Mediterranean countries is wholly captured by the set of "primary" variables considered, the technique studies the complex of relations between the variables so as to analyse their "co-movements" and use these as a basis to construct the so-called "factors" obtained as linear combinations of the initial variables. These factors are nothing other than new variables combining the information originally contained in the basic variables. The analysis produces as many factors as there are primary variables included in the calculation. They can be ordered on the basis of the proportion of total variability accounted for, which can be regarded in turn as a measure of their recapitulative power or rather capacity for synthesis with no significant loss of information.

But what connection is there between the new variables and the original ones? For each "primary" variable considered in the analysis, there is a coefficient (coordinated on the factor) that measures the weight it assumes in the determination of a generic "secondary" variable. The latter is thus characterised to a greater degree by the variables with larger coefficients in terms of absolute value. It is precisely these variables that endow the factor with significance and pinpoint how the individual aspects are connected in the definition of child welfare. Analysis of these coordinates in terms of the different factors identified thus makes it possible to understand the links between the original variables, the study of which then makes it possible to capture the significance of the "secondary" variable. In the terminology of the analysis in question, this is called "baptising" the factor considered.

The task does not end, however, with identification of all the factorial axes. Efforts must instead focus at this point on identifying out of all the factors those that represent welfare.

In actual fact, even if a factor captures a significant proportion of total variability, it could give "preponderant" weight to indicators and relations regarded by the investigator as "marginal" or even irrelevant with respect to the phenomenon examined. Sensitivity must be brought into play to identify out of all the relations present in the factors only those of relevance for the multi-dimensional representation of the phenomenon.

Once the relevant axes have been selected, the analysis also provides the coordinates on the factors for each of the survey units (countries), which can be used for classification purposes either directly or through further processing.

The most immediate way of developing a classification on the basis of a single factorial axis is by regarding the coordinate of each country on that factor as a value of the index. Once the "direction" of the axis has been established with respect to the phenomenon addressed (e.g. a positive coordinate is equivalent to greater comparative welfare and vice versa), it will only be a matter of standardising the values of the coordinates to ensure that each one moves in a defined interval of variation with a known order of magnitude.

Principal components analysis makes it possible not only to develop a general classification of child welfare but also to extend the investigation in other directions. In particular, it is possible to break down the results obtained and identify areas of hardship and areas of success for each country.

The methodology employed identifies the combination of characteristics needed to achieve child welfare and provides the value of the countries' coordinates on every relevant factor. Examination of a country's performance with respect to the elements needed to ensure success in terms of general welfare makes it possible to identify the individual characteristics responsible for that country's position at a certain level in the classification. By maintaining the grouping of indicators by thematic areas, it is also possible to ascertain whether a disappointing result in terms of general welfare should be attributed to shortcomings concentrated in a specific thematic area, e.g. progress in terms of education rather than economic conditions or other elements. On the basis of this information, the analysis can thus be used to suggest provisions or courses of action in order to improve the ranking and hence the level of welfare.

Principal components analysis identifies the characteristics of a phenomenon that presents itself with the same characteristics for all the survey units. Underpinning the analysis is the assumption that there exists a single idea of child welfare and that this is shared by all the countries involved. Despite this, a possible development of the study consists in investigating the data set so as to identify different models of child welfare. The idea is to ascertain whether, in the sphere of the relevant variables, the different countries have adopted different models on the basis of differences in cultural matrix, policy, or lack of resources. This analysis can make it possible to develop as many classifications as there are types of countries identified. This is how we sought to address the problem stated at the end of section 2.

4. CHILD WELFARE IN THE MEDITERRANEAN COUNTRIES

The methodology described above was used to develop an indicator of child welfare in the Mediterranean countries.
Consisting primarily of data from UN sources, the date set of reference was provided by Lynkeus and regards a total of 33 countries belonging to five different geographical areas, namely Arab countries, the Middle East, North Africa, East Europe and Mediterranean Europe. The 39 complete variables for which surveys on each of the countries are available are divided into the six thematic areas of demography, nutrition, health, education, economic indicators and social indicators. Complete lists of the countries and the variables available are provided in tables 1a and 1b.

Table 1a. The Mediterranean countries

Arab countries	1) Bahrain 2) Iran 3) Iraq 4) Kuwait 5) Oman 6) Qatar 7) Saudi Arabia 8) United Arab Emirates 9) Yemen	**Middle East**	1) Israel 2) Jordan 3) Lebanon 4) Syria	**North Africa**	1) Algeria 2) Egypt 3) Libya 4) Morocco 5) Tunisia
East Europe	1) Albania 2) Bosnia 3) Bulgaria 4) Croatia 5) Romania 6) Serbia and Montenegro 7) Slovenia 8) Macedonia 9) Turkey	**Mediterranean Europe**	1) Cyprus 2) France 3) Greece 4) Italy 5) Malta 6) Spain		

Before examining the results, it should be specified that the original data set was adapted to the requirements of the analysis in two respects. Firstly, each of the variables observed was suitably processed so as to make the intervals of variation as homogeneous as possible. This transformation is necessary to ensure that the results of the analysis are not distorted by the different units of measurement used to survey the variables. The aim is to avoid attributing importance to the variables solely because they are characterised by a comparatively greater order of magnitude, as in the case, for example, of GDP per capita and the infant mortality rate.

Table 1b. Complete list of variables
DEMOGRAPHIC INDICATORS
1) Population 0-4
2) Population 0-17
3) Population 5-9
4) Population 10-14
5) Population 15-19
6) Total population
7) Child dependency ratio
8) Birth rate
9) Total fertility rate
10) Infant mortality rate
11) Life expectancy at birth (Total)
12) Urban population
13) Growth of urban population
NUTRITION INDICATORS
1) Underweight births
HEALTH INDICATORS
1) Maternal mortality rate
2) Children aged 1 vaccinated against DPT
3) Children aged 1 vaccinated against polio
4) Children aged 1 vaccinated against measles
5) Number of physicians
6) Total health expenditure
7) Health expenditure per capita
8) Private health expenditure
9) Public health expenditure
10) Out-of-pocket health expenditure
EDUCATION INDICATORS
1) Early childhood care and education (Total)
2) Secondary education (gross)
3) Pre-primary education (gross)
4) Pupil/teacher ratio in primary education
5) Education indicator
ECONOMIC INDICATORS
1) Gross national income per capita
SOCIAL INDICATORS
1) Number of radios
2) Number of televisions
3) Telephone mainlines
4) Cellular mobile subscribers
5) Number of computers in use
6) Internet users
7) Human development index
8) Human development index
9) Population with access to adequate sanitation facilities

The mean value of about ten thousand dollars registered for the former is in fact decidedly out of proportion with respect to the latter, which, being a percentage, is necessarily between 0 and 100.
Secondly, the number of the variables considered was further reduced by eliminating a whole series of indicators regarded as redundant. The indicators in question capture only specific aspects of one and the same phenomenon and therefore add no important information for our purposes. Examples include various social indicators such as the number of television sets, radios and telephones. All three give an indication of the availability of telecommunications, present a high degree of correlation, and do not differ substantially between countries. For these reasons they were combined to form a single indicator labelled availability of telecommunications (TLC). Other indicators were also combined, transformed or eliminated from the analysis on the basis of analogous considerations. Table 2 presents the definitive list of the 19 variables taken into consideration.

The data set thus established constituted the basis for the application of principal components analysis (PCA).
As pointed out, this makes it possible to provide a concise representation of the phenomenon through the extraction of factorial axes or factors. The first factor accounts for the largest amount of variation while the second, orthogonal to the first, furnishes a representation of the residual variation, and so on. It is therefore clear that the relations described by the first factorial axes are those capturing the highest proportion of information. The capacity for synthesis thus depends on how large a proportion of variability is accounted for by the first or at most the first two factorial axes.
The application of PCA to the 19 variables listed in table 2 made it possible to extract an initial factorial axis accounting for no less than 49 per cent, i.e. just under half, of the total variability. As the capacity for synthesis of the first axis is decidedly high, the relations it describes already proved sufficient in themselves to provide a description of the latent phenomenon to be accounted for. A graphic representation is presented in figure 1.

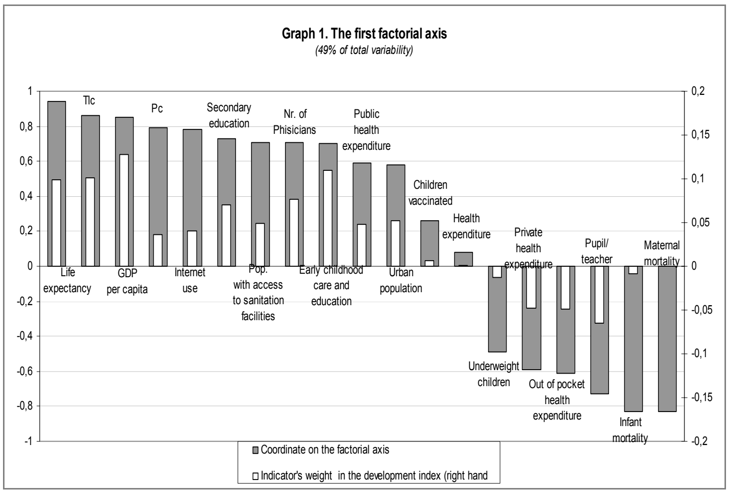
Table 2. List of variables included in the analysis
DEMOGRAPHIC INDICATORS
1) Infant mortality rate 2) Life expectancy at birth (Total) 3) Urban population
NUTRITION INDICATORS
1) Underweight births
NEALTH INDICATORS
1) Maternal mortality rate 2) Children vaccinated 3) Number of physicians 4) Health expenditure as % of GDP 5) Private health expenditure (as % of total) 6) Public health expenditure (as % of total) 7) Out-of-pocket health expenditure (as % of total)
EDUCATION INDICATORS
1) Pupil/teacher ratio in primary education 2) Early childhood care and education (Total) 3) Secondary education (gross)
ECONOMIC INDICATORS
1) Gross national income per capita (index)
SOCIAL INDICATORS
1) Availability of TLC 2) Number of computers in use 3) Internet users 4) Population with access to adequate sanitation facilities

4.1 Factorial analysis

The coordinates of the individual variables on the first axis (or factor) are shown in the diagram with the dark histogram. Each coordinate is to be read on the basis of two different criteria. The first is related to the scale of the coordinate in terms of absolute value (i.e. in graphical terms

to the “largeness” of the histogram). It registers the intensity of the link between the underlying variable and the phenomenon to be explained. High coordinates (such as life expectancy, which has a coordinate equal to approximately 0.9) are therefore comparatively more correlated with child welfare than coordinates close to zero.
The second criterion instead regards the “proximity” or “distance” of the coordinates themselves. Positive coordinates located close to one another mean that the two variables tend to have a high degree of correlation, whereas coordinates that have opposite signs and are therefore “distant” indicate that the underlying variables are characterised by negative correlation. For example, it emerges from the data that where life expectancy is high (a positive and large coordinate) there is also a high degree of availability of telecommunications (TLC, again with a large and positive coordinate and hence “near”). Conversely, the same countries register a very low rate of infant mortality (a variable with a decidedly negative coordinate that is in “opposition” to the previous ones in that it attains the “maximum distance” from the positive coordinates).

In order to facilitate the reading of the results, the coordinates of the first factor were placed in order of decreasing magnitude. In diagram 1, proceeding from left to right, we thus encounter the positive coordinates first and then the negative ones.



The first point to emerge immediately and clearly is that the relations between the variables selected and the underlying phenomenon are consistent with the most intuitive *a priori* expectations, thus confirming that the first axis constitutes a credible synthesis of child welfare.

The analysis stresses in fact the existence of a positive correlation between child welfare and comparatively higher levels of life expectancy. There is also a positive correlation between welfare and level of GDP per capita, the indicator par excellence of economic conditions and also, at least to some extent, of living conditions. Welfare is also linked to a series of indicators of social development, such as the availability of telecommunications, possession of personal computers, and use of the Internet.

To a lesser degree (the coordinates gradually decrease in weight as we proceed toward the right), we can observe that the phenomenon is also characterised by high rates of school attendance, the presence of adequate health services, and the availability of medical assistance.

At the other end of the diagram, we find that child welfare is negatively correlated (the coordinates are in fact negative) with a different set of indicators. Among the variables connected with education, for example, we note that a low ratio of pupils to teachers is an index of welfare. The same holds in the demographic field for low rates of infant and maternal mortality.

The result obtained with respect to indicators of health conditions instead appears to be diversified. First and foremost, it emerges that total health expenditure (as a proportion of GDP) does not significantly characterise this factorial axis, its coordinate being the one closest of all to zero. It is, however, noted that the public component of this expenditure is associated positively with the factors of development, whereas its private components are characterised by negative coordinates). This result reflects the fact that the sample of countries is unbalanced to some degree. The obvious cases of success in terms of child welfare, like France or Italy, are all countries in which a largely public health system has been consolidated for a long time. The countries where welfare is less widespread are instead ones where the health system is still a limited presence or in any case supported only by private capitals. The same operation might not give the same result if repeated at the world level and should therefore be interpreted with all due caution.

Even though the relations identified are somewhat intuitive as a whole, the nature of child welfare as a multidimensional phenomenon is confirmed by the presence on the first factor of a number of variables characterised by high coordinates in terms of absolute values. The multidimensional nature of welfare is adequately represented on the first axis, whose power of synthesis is, as pointed out, very high, and there is thus no need to search the other factors for further elements serving to account for the phenomenon.

4.2 Determinants of the development index

In order to determine the importance of each variable in accounting for child welfare, further steps were undertaken, completing the information provided by the coordinate of the variable on the factor with information about its scale. To this end, it was important to estimate how far policy actions in specific areas of intervention influence child welfare. Translated into statistical terms, this meant estimating the effect on the welfare index of impulses, measured as standard deviations of the variables introduced into the analysis. The contribution of each variable is shown in diagram 1 by the white histograms.

Seen in this way, the picture changes substantially with respect to the view based on the values of the coordinates. As rightly expected, the most important role affecting child welfare is played by income per capita. This predominance is, however, anything but overwhelming in that income determines “only” 12.8% of total variation. These figures lead us away from any purely economistic view.

Closer examination shows in fact that it is all the other elements as a whole that play a predominant role in determining the index of child welfare. Attention should be drawn immediately to the key role played by the variables of education. Particular importance attaches to the indicator of early childhood care and education, which accounts by itself for 11 per cent of total variation. Considered as a whole, the variables connected with education constitute one of the most important areas of intervention with a total weight of 24.4 per cent in the index of child welfare.

The set of indicators associated with health and nutrition also play an important role in defining the index. Their combined weight amounts to 24.3 per cent of the total, which is substantially on a par with that of education. The number of physicians plays a dominant role in this respect, with a weight in the index of 7.6 per cent. Among the other elements considered in the area associated with health conditions, public health expenditure determines 4.8 per cent of the index, whereas the opposite appears to hold for the private component of health expenditure, a decrease in which appears to be associated positively with development. Finally, with respect to the variables connected with nutrition, it should be noted that the percentage of children born underweight plays a very limited part in determining the index (1.5 per cent).

Factors of social development account for another 22.6 per cent of the total. The predominant variable here is the availability of telecommunications, which accounts for 10 per cent of the index. This is followed by another two indicators of technological development, namely the use of personal computers and the Internet, which together make up a contribution of 7.6 per cent. Nearly 5 per cent of the index is instead determined by the availability and quality of sanitary facilities.

The remaining part of the index, about 16 per cent, is determined by demographic factors. Life expectancy is the most important, accounting for 10 per cent of the variation of the index. Following this, we note that the phenomenon of urbanisation, expressed as an increasing percentage of the population living in urban agglomerates, accounts for 5.2 per cent of welfare. The infant mortality rate instead accounts for just under one per cent of the index.

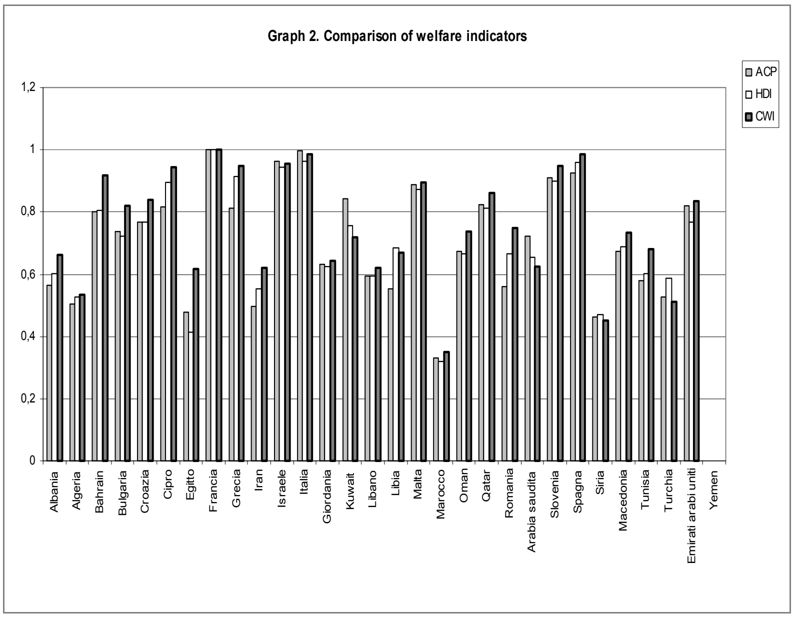
Though important in itself, the contribution of GDP is ultimately not preponderant with respect to other elements in the determination of child welfare. Generally speaking, the result obtained fully justifies the attempts to construct extended indices of welfare that take other determinant factors of development into consideration alongside those of a strictly economic nature.

What emerges in terms of policy indications is the need for action designed not only to stimulate the dynamics of production but also to guarantee the population improved levels of education and health so as to accompany economic growth with the creation of a society that does not lack the basic elements for survival with dignity and is ready to tackle the challenges of technological development. At the statistical level, the analysis suggests the weighting and proportions of such operations.

4.3 Child welfare in the different countries

The coordinates and the weights of the variables were used finally to construct a classification of the different countries as regards child welfare.

Diagram 2 compares the value of the welfare index obtained on the basis of PCA with the values of the human development index and child welfare index (as published in the article by Van der Gaag respectively on pages 63 and 65).[3]



The index developed on the basis of PCA moves in line with the other two indicators, the degree of correlation being over 97 per cent with the human development index and about 95 per cent with the child welfare indicator.

The classifications drawn up on the basis of these values are shown in table 3. It is observed that the index based on PCA identifies the same cases of success (for example, France and Italy are again placed first and second respectively).

[3] For the purposes of this comparison, the values of the three indices were normalised and framed in a common field of variation.

Table 3. Comparison of classifications						
	Welfare index based on ACP		Human Development Index (HDI)		Indice di benessere del bambino (CWI)	
	Value	Ranking	Value	Ranking	Value	Ranking
France	1,000	**1**	1,000	**1**	1,000	**1**
Italy	0,997	**2**	0,965	**2**	0,987	**2**
Israel	0,964	**3**	0,943	**4**	0,956	**4**
Spain	0,927	**4**	0,959	**3**	0,986	**3**
Slovenia	0,909	**5**	0,900	**6**	0,949	**5**
Malta	0,889	**6**	0,873	**8**	0,896	**9**
Kuwait	0,844	**7**	0,754	**13**	0,717	**17**
Qatar	0,824	**8**	0,814	**9**	0,861	**10**
United Arab Emirates	0,820	**9**	0,767	**11**	0,836	**12**
Cyprus	0,818	**10**	0,894	**7**	0,945	**7**
Greece	0,813	**11**	0,912	**5**	0,946	**6**
Bahrain	0,802	**12**	0,804	**10**	0,919	**8**
Croatia	0,767	**13**	0,766	**12**	0,837	**11**
Bulgaria	0,737	**14**	0,721	**14**	0,818	**13**
Saudi Arabia	0,722	**15**	0,654	**19**	0,624	**22**
Oman	0,674	**16**	0,665	**18**	0,738	**15**
Macedonia	0,673	**17**	0,690	**15**	0,735	**16**
Jordan	0,633	**18**	0,625	**20**	0,643	**21**
Lebanon	0,596	**19**	0,593	**23**	0,621	**23**
Tunisia	0,581	**20**	0,602	**21**	0,679	**18**
Albania	0,563	**21**	0,602	**22**	0,662	**20**
Romania	0,560	**22**	0,666	**17**	0,749	**14**
Libya	0,555	**23**	0,686	**16**	0,668	**19**
Turkey	0,525	**24**	0,586	**24**	0,510	**27**
Algeria	0,504	**25**	0,527	**26**	0,532	**26**
Iran	0,495	**26**	0,553	**25**	0,621	**24**
Egypt	0,479	**27**	0,413	**28**	0,616	**25**
Syria	0,462	**28**	0,470	**27**	0,452	**28**
Morocco	0,330	**29**	0,321	**29**	0,351	**29**
Yemen	0,000	**30**	0,000	**30**	0,000	**30**

As regards failure, the three classifications again place the same countries (Morocco and Yemen) in the bottom and next to bottom positions.

The marked degree of analogy with the other synthetic indicators leads us to regard the results obtained as plausible and hence also of importance with respect to policy decisions, which analyses of this type can help to focus more sharply by weighting the critical areas in which action would prove most effective.

METHODOLOGICAL APPENDIX

Factorial Analysis

Factorial analysis encompasses a set of statistical techniques of the multidimensional type that seek to condense the information of a large mass of interconnected data, pursuing the objective of maintaining the highest possible explanatory power of the information while simplifying its content and making it more accessible.

The approach in factorial analysis associated with the French school of *Analyse des donnés* is based on a formalization of the data matrix on a non-probabilized space (especially on a Euclidean vectorial space) and can be summarized by a model of the type

$$(\underline{D}, R^m, (R^n), \underline{M}),$$

where R^m is the vectorial space of the units (R^n the vectorial space of the variables) and $\underline{M}$ a metric on that space.

The purpose of this model is the simultaneous analysis of a range of variables representing the matrix $\underline{D}$ as a cluster of points in the R^m space or the R^n space.
It proves difficult for fairly large m or n (as in the case in point) to offer an interpretation of the data through a direct reading of the positions of the points in spaces of such dimensions A statistical model thus framed must necessarily have recourse to *subspaces* of lesser dimension on which to project the points so as to interpret the distances between them as well as the overall picture of their reciprocal associations and positions.

The key problem is thus to determine these subspaces of projection so that the reduction in dimension preserves the original information to the highest possible degree and to reduce the distorting effect that the operation of projection has on the real distances between the points. In other words, it is a question of identifying the subspace that minimizes such distortion and fits the observed situation better than any of the other subspaces parallel to it.

The problem is solved through identification of the *factorial axes or dimensions*.[4] These constitute new statistical variables (obtained as linear combinations of the original variables) that reduce and condense the matrix D so as to offer a reading of the original information that is simplified but possibly richer in meaning.

In addition to the creation of reduced dimensions through which to obtain a simple and significant reading of the information contained in complex data matrices, an essential role is played by the following primary characteristics of the approach described:

The duality between units and variables is made explicit, thus permitting valorization of the dimension of the statistical units.
It is possible to seek out forms and structures that are hard to identify a priori.
Systematic use is made of graphic representations, which prove very useful and simple in terms of interpretation and communication.

Being orthogonal constructions, the factorial axes prove *informative* in that each axis explains a part of the overall variability of the phenomenon that is not accounted for by the other axes.
Moreover, these axes are identified in decreasing hierarchical order so that the variability explained by the first is greater than that explained by the second, which is in turn greater than that of the third, and so on.

---

[4] For further details of this approach, due primarily to the French school and better known as *Analyse des données*, see the following:
- J.P.Benzecri, L'Analyse des Données, Tome 1: La Taxinomie, Tome 2: L'Analyse des corrispondances, Dunod, Paris, 1973;
- R.Coppi, *Appunti di statistica metodologica:analisi lineare dei dati*, Dipartimento di Statistica e Calcolo delle Probabilità, Università degli Studi "La Sapienza", Rome, 1986;
- L.Lebart, A.Morineau, J.P.Fenelon, *Traitement des données statistiques*, Dunod, Paris, 1982

Finally, interpretation of the results of the analysis is based on various indices of the quality of representation, differing in relation to of the specific technique used but always constructed on the basis of the factorial axes.

Principal Components Analysis

Principal components analysis is a multivariate statistical technique used in the sphere of data synthesis and in the study of latent variables.
It often happens that a large number of variables are available for the phenomenon under examination, each of which representing one or more aspects of it. In such conditions it is clearly necessary to represent the phenomenon by means of a smaller number of variables derived from those originally observed and preserving as much information as possible.
The principal components method solves this problem by constructing a set of p "artificial" variables ($\xi_1, \xi_2, ..., \xi_p$) obtained as a linear combination of the k original variables ($X_1, X_2; ....X_k$) with $p < k$, these being uncorrelated and such that each of them has maximum variance.
That happens because a statistical variable with a high degree of variability tends to assume modalities differing greatly from one another. If reference is thus made to a collection of individuals on which the statistical variable is based, its high degree of variability means that the individuals tend to differ markedly in terms of the variable question.
In other words, it is necessary to make sure that the artificial variable is capable of providing a significant informative contribution, or rather that the principal components $\xi_1, \xi_2, ..., \xi_p$ are capable of summarizing or synthesizing the phenomenon under examination without this involving any important loss of information.
The purpose of principal components analysis in the perspective outlined above is to represent the units and variables in subspaces of low dimension (contained respectively in R^m and R^n) that preserve to the greatest possible degree the statistical information presented by the two clusters of points present in these spaces.

A statistical variable $X = (X_1, X_2, ..., X_k)'$ is considered and its vector of means and its matrix of variances and covariances are indicated respectively with μ and Σ.
As stated above, our aim is to construct an artificial variable ξ_1 as a linear combination of the variables originally observed:
$\xi_1 = \alpha_{11} X_1 + \alpha_{12} X_2 + ... + \alpha_{1k} X_k = \alpha_1' X$
such that its variance proves maximum.
Since $E(\xi_1) = \alpha_1' \mu$ e $Var(\xi_1) = \alpha_1' \Sigma \alpha_1'$, it is evident that identification the first principal component will entail determining the value of the vector with k components α_1.
In order to obtain an unambiguously determined solution, it is necessary to impose the following constraint on the α_1 vector components:
$\alpha_1' \alpha_1 = 1$
This is essentially a question of solving the following bound minimum problem:

$$\begin{cases} \max \quad \alpha_1' \Sigma \alpha_1 \\ \quad \alpha_1' \alpha_1 = 1 \end{cases}$$

By means of the method of Lagrange multipliers, we obtain the condition
$\Sigma \alpha_1 = \lambda \alpha_1$
where λ is the Lagrange multiplier.
The above is an equation of autovalues and shows how λ is an autovalue of the matrix Σ while α_1 is the normally unitary corresponding autovector.
At the same time, the equality $\Sigma \alpha_1 = \lambda \alpha_1$ and the constraint $\alpha_1' \alpha_1 = 1$ entail that

$$Var(\xi_1) = \alpha_1' \Sigma \alpha_1 = \lambda \alpha_1' \alpha_1 = \lambda$$

As a result, λ can only be the maximum autovalue of the matrix Σ, which we shall denote as λ_1, and α_1 is the normally unitary autovector corresponding to it.
We have thus defined the variable

$$\xi_1 = \alpha_1' X$$

which we shall call the first principal component.
The other principal components can be calculated in the same way. For example the second principal component ξ_2 is again defined as a linear combination of the original variables

$$\xi_2 = \alpha_{21} X_1 + \alpha_{22} X_2 + ... + \alpha_{2k} X_k = \alpha_2' X$$

and is characterized by the following properties:
1. It is not correlated with ξ_1.
2. It has the highest variance of all the variables not correlated with ξ_1.
The second principal component is calculated by solving the bound maximum problem

$$\begin{cases} \max \quad \alpha_2' \Sigma \alpha_2 \\ \alpha_2' \alpha_2 = 1 \\ \alpha_2' \Sigma \alpha_1 = 0 \end{cases}$$

where the second constraint derives from the cancelling out of the covariance between ξ_1 and ξ_2. As before, use of the method of the Lagrange multipliers provides the equation of autovalues
$\Sigma \alpha_2 = \lambda \alpha_2$
and thus, as before, λ cannot be other than the second largest autovalue of the matrix Σ while α_2 is the normally unitary corresponding autovector.
The other principal components can be defined in the same way. It is important to note that if the variables $X_1, X_2; ....X_k$ are uncorrelated (the matrix of the variances and covariances is diagonal), they are exactly equal to the principal components.
If the experience and sensitivity of the investigator unquestionably play a predominant role in the interpretation of the principal components, it is also clear that the generic coefficient α_{jh} of the linear combination corresponding to the variable X_h measures the weight assumed by that variable in the determination of the principal component ξ_j.

This means that the principal component ξ_j will be characterized to a greater degree by the variables X_h with the largest coefficients α_{jh} in terms of absolute value. It is thus precisely these variables that give significance to the principal component.

As mentioned above, principal components analysis is often used in the study of "latent" variables, i.e. variables that are not directly observable. In this connection, the artificial variables created by the analysis represent an attempt to observe and measure variables that are hidden and not susceptible of direct exploration.

In any case, whatever the reason for recourse to principal components analysis, one of the problems to be tackled regards the number of principal components taken into consideration.

To this end, the following method is most often used:

1. Set a threshold $I^* =$ "proportion of total variance to be explained".
2. Select the smallest number p of principal components such that $I_p \geq I^*$.

It is obvious that the proportion of total variance explained by the first p principal components can be regarded as a measure of their "explanatory power". A high value of I_p indicates that the first p principal components account for a high proportion of the total variance.