Exploring competitiveness and wellbeing in Italy by spatial principal component analysis

Carlo Cusatelli, Massimiliano Giacalone, Eugenia Nissi

1. Introduction

The statistical observation of a complex social phenomenon such as of *well-being* includes a series of logical operations that lead to empirical indicators suitable for its study. The so-called statistical operationalization, which in this case will include the following basic steps:
- definition of the complex concept of well-being, which cannot be measured directly;
- specification of the concept, or its decomposition into directly observable and measurable dimensions and sub-dimensions;
- choice of indicators, measuring instruments of each dimension and sub-dimension.

It is not easy to define the concept of well-being: it concerns a social phenomenon with very different semantic and interpretative orientations that have their roots in various disciplines, such as Economics, Sociology, Psychology, Urban Planning and others. A concept that embraces all aspects of living can therefore only be part of a theoretical model built ad hoc. To begin to evaluate the problems that led to the construction of the model used in this work, we want to briefly recall the excursus that over time has led to the different formulations of an evolving concept: *quality of life*.

Some authors[1] identify the birth of the problem in question in the period of the industrial revolution. At that time, the profound changes in the living conditions of the European population led some researchers[2] to study, together with the income earned by the new types of workers, the price of a basket of goods consumed by them and the composition of their expenses, identifying significant differences between the structure of the budgets of the less well-off classes (in which almost all of the income was consumed for food) and that of the wealthier classes (who saw the relative share of food expenditure diminish). Such research, of a purely descriptive nature, spread during the nineteenth century in various European countries[3] and, essentially linked to the study of the subsistence budgets of working families, maintained a purely economic significance of studies on the level of survival of these families.

Subsequently, together with the evolution of the economic conditions of the working families, the concept under examination took on a different meaning, becoming more and more identified with that of satisfying needs that are no longer exclusively nutritional. One of the terms relating to this evolving phenomenon was that of *standard of living*, by which some scholars[4] defined the set of goods and services used in families whose type of life (*mode of life*) was determined by different parameters and social characteristics. According to a different

[1] See, e.g., W.S. and E.S. Woytinsky, *World population and production. Trends and outlook,* The Twentieth Century Fund, New York, 1953; J. Fourastié, *Machinisme et bien-être, niveau de vie et genre de vie en France de 1700 à nos jours,* Les éditions de minuit, Paris, 1962.
[2] See, e.g., F. Le Play, *Les ouvriers européens; études sur les travaux, la vie domestique et la condition morale des populations ouvriéres de l'Europe, précédées d'un exposé de la méthode d'observation*, Paris, 1855.
[3] See, e.g., E. Engel, *La consommation comme mesure du bien-être des individus, des familles, des nations*, in Bulletin de l'Institut International de Statistique, Tome II, Roma, 1887.
[4] See, e.g., A. Bowley, *The nature and the purposes of the measurement of social phenomena*, P.S. King & Son Ltd., London, 1923.

Carlo Cusatelli, University of Bari Aldo Moro, Italy, carlo.cusatelli@uniba.it, 0000-0003-3770-3060
Massimiliano Giacalone, University of Naples Federico II, Italy, maxgiacit@yahoo.it, 0000-0002-4284-520X
Eugenia Nissi, Gabriele d'Annunzio University, Italy, nissi@unich.it, 0000-0003-3440-601X

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

approach, other authors[5] addressed themselves towards measuring the individual psychic satisfactions that income could provide through the consumption of goods and services, formulating the utilitarian concept of *niveau de confort*. Still others[6], considering a broader meaning of well-being, came to the concept of a *level of living* which, by loosening the centrality of economic needs, also penetrated into the social areas of *welfare*.

Totally emancipated from definitions of an economic matrix that are not suited to an approach inspired by components of a purely social nature, the concept of *quality of life* dates back to the early Seventies, a period in which a particular and active political-social climate was developing. as opposed to the prevailing economist conception of progress: in Western societies characterized by high rates of economic growth (the *most*, the quantity), it was beginning to doubt that such growth could always be equivalent to social progress (the *best*, the quality). Roads choked by traffic, air pollution, difficult accessibility to services theoretically aimed at all citizens, the spread of new forms of poverty, difficulties in interpersonal relationships are just some of the phenomena that are easy to find in the most industrialized and, particularly, in urbanized contexts, where wealth and population are concentrated but also inequality and social hardship.

Another fundamental guideline of the debate on the quality of life was the requirement that the concept also contemplate the subjective aspects of human existence: the term *quality* implies in fact a personal judgment that is generally not measurable except through subjective indicators. Through the latter it is in fact possible to grasp the internalization of social problems by individuals (attitudes, judgments, perceptions, concerns, etc.). Furthermore, subjective indicators make it possible to complete and specify the information collected by means of objective indicators regarding the aspects (material and otherwise) of the quality of life, towards which individuals perceive a different satisfaction.

Given the interdisciplinary value of the phenomenon, and lacking a unanimous definition of the concept of well-being, the theoretical model formulated in this work stems from the consideration that the healthy evolution of the territory must correspond to a trend towards economic growth that also brings with it another type prosperity: efficiency and effectiveness of services, interpersonal relationships, culture, good housing conditions, and so on. But if on the one hand urbanization offers such undoubted localization advantages, on the other it produces growing disadvantages; therefore the individual who expects high living standards from it may, on the contrary, find himself paying the price attributable to the malaise produced by the degradation: atmospheric pollution, crime, etc.

Assuming that the urbanized one is still the space in which these expectations can be realized more easily, it becomes inevitable to ask the question of its livability when planning and governing the modernization processes of today's society. Therefore, in order to study the phenomenon in question in relation to a more circumscribed collective than that of the entire population of a nation (to which most of the existing studies on the subject refer for reasons of international comparability), it is advisable to examine a territorial area of great socio-demographic importance: the provincial one. The statistical units of this research are therefore represented by the 107 Italian provinces with respect to the objective indicators that will be described later.

The insights expressed so far highlight the evanescence of the concept of well-being which must therefore be stopped in its empirically measurable dimensions. Well-being is an important element as it discriminates against human aggregates by enriching the image of a territory, strengthening its attractiveness, highlighting, in short, its state of health, that is the ability to fulfill its different roles: it is a place of private residence, social place where meetings and

[5] See, e.g., Bureau International du Travail (BIT), *Les méthodes d'enquête sur les budgets familiaux*, Etudes et Documents, Série N, n.9, Genève, 1926.

[6] See, e.g., D.E. Christian, *International social indicators: the OECD experience*, Social Indicators Research, 1974, n °1, p. 169-186.

interactions are easier, public place where collective demand services can be used to the extent that agglomeration externalities exist. These are precisely the guidelines on which the model adopted here to evaluate the problem of quality of life is based: an integral concept of human development capable of grasping, in addition to the economic one, also other aspirations.

Nowadays a significant amount of data and information is available on topics of local interest which, especially at an objective level, can lead to a fairly complete examination of well-being according to the directions just outlined. Social concerns often examined in order to identify and compare levels of well-being between nations constitute a useful reference in the selection of thematic areas and indicators that can be adopted for this study. The topics that most interest the sphere of well-being can be summarized in the following areas of investigation: environment, housing conditions, roads, work, public services, crime, cultural level.

Within the aforementioned areas, suitable means of indication must therefore be sought. Their identification constitutes a very important phase of this study: after having defined and specified the object of the research and identified the statistical units of reference (the Italian provinces), it is necessary to select the indicators that are best able to measure the various aspects of the well-being.

Once the indicators were chosen on the basis of the foregoing, the data collection often provided only the raw material of the final product represented by the measuring instruments of well-being. The next step therefore concerned the construction of the indicators: this phase consists of all those operations in which some initial data are weighted or variously combined with each other in order to make them statistically comparable and theoretically representative of the phenomenon to be studied.

The simplest family of indicators is that of the so-called primary measures: they concern the amount of individual characters possessed by each statistical unit. At a higher level of complexity are the simple (or elementary) weighted indicators, constructed by dividing the primary measure by a reference variable (which is often another primary measure) called the basic measure; this operation, eliminating the source of variation determined by the basic measure, has the purpose of legitimizing the comparability of the data relating to the various statistical units.

Often, then, the need arises to combine the different simple indicators into compound (or synthetic) indicators, especially when the relationship between phenomenon and elementary indicator is not simply one-to-one, but rather problematic and complex: the same phenomenon can in fact be measured by means of several simple indicators, all different but sometimes partial with respect to the final dimension to be represented. They can be integrated into one or more models that give as an answer the level of each constitutive aspect of the phenomenon considered (partial compound indicators), or its overall level (a global compound indicator). These aggregations also make it possible to better visualize the state conditions, especially when one wishes to make comparisons between different realities, which are also necessary in this research so that the representation of the livability of our country can be interpreted.

The work is organized as follows: in section two some methodological aspects related to principal component analysis for spatial data will be presented; finally, an application to the the data of BES at local level NUTS 3 will be presented.

2. Spatial Principal Component Analysis

Principal Component Analysis (PCA) is one of the most popular multivariate statistical technique used for reducing data with many dimensions, and often wellbeing indicators are obtained using PCA: it is implicitly based on a reflective measurement model that it non suitable for all types of indicators. Mazziotta and Pareto (2013) in their paper discuss the use and misuse of PCA for measuring well-being. The classical PCA is not suitable for data collected on the territory because it does not take into account the spatial autocorrelation present in the data.

Spatial PCA techniques, specifically designed for spatial effects are available. The Geographically Weighted Principal Component (GWPC) is a method that adapt PCA for spatial effects. Given *n* observation x_i it depends from its location *i* in the space with coordinates (u, v) such that supposing $x_i \sim \mathcal{N}(\mu(u,v), \Sigma(u,v)$, the Geographically Weighted Principal Components are obtained through the decomposition of geographically weighted variance-covariance matrix (Harris et al 2011):

$$\Sigma(u,v) = \boldsymbol{X}^t \boldsymbol{W}(u,v) \boldsymbol{X}$$

where $\boldsymbol{W}(u,v)$ is a diagonal matrix of weights. Different kernels functions can be employed to generate the diagonal matrix of weights. Hence the local principal component at location (u_i, v_i).

$$\Sigma(u_i,v_i) = \boldsymbol{L}(u_i,v_i) \boldsymbol{V}(u_i,v_i) \boldsymbol{L}(u_i,v_i)^T$$

where $\boldsymbol{L}(u_i,v_i)$ is the matrix of geographically weighted eigenvectors and $\boldsymbol{V}(u_i,v_i)$ is the diagonal matrix of the geographically weighted eigenvalues.

Considering a set of *p* variables, the GWPC provides *p* components, *p* eigenvalues, *p* set of component loadings and *p* sets of component scores for each location in the study area.

An alternative way to assess the spatial variability of data in PCA is to consider the Locally Weighted Principal Component (LWPC) applied to the situation when data are not described well by an universal set of principal component (Tipping and Bishop 1999,). This technique use a moving window weighting approach in the data space. For each individual LWPCA around *x*, neighboring data point are first weighted according to some distance decay kernel function. Each observation is then multiplied by its respective weight and a standard PCA algorithm is (locally) applied to this weighted data.

Spatial effects can also be taken into account when PCA is combined with a measure of spatial autocorrelation. Jombart et al. (2008) have introduced a modification of PCA (called sPCA) to investigate the pattern of spatial variability of multivariate spatial pattern. The presence of spatial autocorrelation is measured using Moran's I (Moran, 1950). sPCA provides PC scores that summarize both the aspatial variability and the spatial autocorrelation structure in geographical space.

3. Empirical Results

The application consider the data of BES at local level NUTS 3, a system of equitable and sustainable well-being indicators at small-regions level that are consistent with the national Bes measures. To meet the statistical information needs of local communities, Istat designed Bes at local level in cooperation with local authorities, investigating the specific information needs of Italian Municipalities, Provinces and Metropolitan Cities and tuning a shared theoretical framework. Bes measures at local level maintain a high level of quality and consistency with the Bes indicators system and constantly follow the evolution of the Bes framework. The two frameworks share a core of common and harmonized indicators. In addition, Bes at local level includes specific well-being indicators, concerning some issues that are related to responsibilities and functions of local authorities (Istat, 2020).

The set of indicators, illustrating the 12 domains relevant for the measurement of well-being, is updated and illustrated annually in the Bes report. In 2020, the set of indicators has been expanded to 152 (it was 130 in previous editions), with a deep revision that takes into account the transformations that have characterised Italian society in the last decade, including those linked to the spread of the COVID-19 pandemic.

The first step in spatial analysis is to asses if a source of spatial correlation is present in the

data or not. Among all the possible alternatives, Moran's *I* is the test used for assessing the presence of spatial autocorrelation. The results of Moran's *I* test reveal a statistically positive spatial autocorrelation for each pillar (Tab. 1).

Table 1. Moran's *I* test of spatial autocorrelation.

Indicator	Statistics	p-value
Life expectancy at birth (male)	0.3481	$6.064e^{-08}$***
Life expectancy at birth (female)	0.3891	1.542^{-09}***
Infant mortality rate	0.1827	0.002119***
Mortality rate for road accidents (15-34 years old)	-0.01504	0.5337
Age-standardised cancer mortality rate (19-64 years old)	0.2299	0.0001871***
Age-standardised mortality rate for dementia and related illnesses (people aged 65 and over)	0.2715	$1.568e^{-05}$***
Participation primary school	0.0508	0.1858
Participation in upper secondary education	0.2683	$1.944e^{-05}$***
Participation in tertiary education (19-25years old)	0.1767	0.0028***
Early leavers from education and training	0.5551	$2.2e^{-16}$***
Young people who do not work and do not study	0.5509	$2.2e^{-16}$***
Level of literacy	0.4651	$8.681e^{-13}$***
Level of numeracy	0.6191	$2.2e^{-16}$***
Employment rate of people 20-64 years old	0.4766	$3.406e^{-13}$***
Non-participation rate (15-74 years old)	0.4836	$1.39e^{-13}$***
Incidence rate of fatal occupational injuries or injuries leading to permanent disability	-0.0755	0.8381
Employment rate of women with and without children	0.1601	0.005711**
Per capita adjusted disposable income	0.3995	$7.233e^{-10}$***
Distribution of IRPEF incomes	0.4753	$4.056e^{-13}$***
Quality of dwellings	0.0622	0.1392
Number of people in workless households	0.5518	$2.2e^{-16}$***
Households with suffering bank debts	0.3152	$6.986e^{-07}$***
Volunteers in no-profit organizations (per 100 residents aged 14+)	0.2663	$1.202e^{-05}$***
No-profit organizations	0.3142	$7.304e^{-07}$***
Social cooperatives	0.3234	$3.047e^{-07}$***
of paid workers in local units of social cooperatives	0.2441	$7.753e^{-05}$***
Homicide rate	0.0589	0.1454
Burglaries	0.3746	$6.744e^{-09}$***
Pickpocketing	0.0496	0.1843
Robberies	-0.0542	0.7635
Presence of historic rural landscapes	-1.072	0.5659
Libraries	0.3598	1.271^{-08}***
Museums	0.3183	8.222^{-08}***
Visitors of libraries	0.2037	0.0003068***
Visitors of museums and similar institutions	-0.0427	0.7128
Drinkable water supplied every day per capita	0.4138	$1.651e^{-10}$***
Exceeding of the daily limit for the protection of human health for PM10 (Maximum number)	0.2217	0.000314
Urban parks and gardens	0.4735	$4.718e^{-13}$***
Protected Natural Areas	0.0242	0.307
Urban green areas	0.0386	0.1728
Noise pollution	-0.0283	0.628

The results shows that spatial dimension is relevant for all sustainable well-being dimensions and within each dimension, for most indicators. This evidence gives other further elements in favour of the application of Spatial PCA (sPCA) than a classical PCA in investigating determinants of provinces well-being for construction spatial composite indicator.

There is a strong spatial differentiation between positive and negative spatial correlation for each principal component obtained, recorded in the North and the South part of the country, respectively.

The spatial nature of sPCA ensure that the percentage of total variance explained can be decompose into pure variability and spatial autocorrelation.

The spatial patterns in the proportion of the explained variance vary significantly across the studied region, allowing in such a way to highlight territorial urban differences. In general, for the majority of the urban well-being domains the highest PTVs (Proportion of the Total Variance) are located in the Province capital cities in the south of Italy.

For the domain Health, the thematic maps of the local principal components reveal the

presence of a global spatial pattern.

By mapping the spatial variation of the first local component of the Education domain, we find out that it mainly considers the participation to primary school and this elementary indicator dominates in most urban areas, with some exceptions for a number of Province capital cities of Calabria and Sicily, where the leading variable is represented by the early leavers from education and training, aimed to capture the problem of school dropout. The dimension Education is characterized by local and global pattern.

For work and life balance pillar we can observe a lesser geographic variation in the influence of each variable on the first component: for the majority of cities the dominant variable is the employment rate of women.

References

Allardt, E. (1981). Experiences from the comparative Scandinavian study, with a bibliography of the project. *European Journal of Political Research*, **9**, pp. 101-111.

Andrews, F., Szalai A. (1980). *Quality of life: comparative studies*, Sage, London, (UK).

Bowley, A. (1923). *The nature and the purposes of the measurement of social phenomena*. P.S. King & Son Ltd., London, (UK).

Bureau International du Travail – BIT (1926). *Les méthodes d'enquête sur les budgets familiaux*, Etudes et Documents, Série N, **9**, Genève, (CH).

Christian, D.E. (1974). International social indicators: the OECD experience. *Social Indicators Research*, 1974, **1**, pp. 169-186.

Engel, E. (1887). La consommation comme mesure du bien-être des individus, des familles, des nations, in *Bulletin de l'Institut International de Statistique*, Tome II, Héritiers Botta, Roma, (IT).

Fourastié, J. (1962). *Machinisme et bien-être, niveau de vie et genre de vie en France de 1700 à nos jours*. Les éditions de minuit, Paris, (FR).

Jombart, T., Devillard, S., Dufour, A. B., Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**(1), pp. 92-103.

Harris, P., Brunsdon, C., Charlton, M. (2011). Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, **25**(10), pp. 1717-1736.

ISTAT(2020), Rapporto BES.

Le Play, F. (1855). *Les ouvriers européens; études sur les travaux, la vie domestique et la condition morale des populations ouvriéres de l'Europe, précédées d'un exposé de la méthode d'observation*, Imprimerie Impériale, Paris, (FR).

Mazziotta, M., Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one. *Rivista Italiana di Economia Demografia e Statistica*, **67**(2), pp. 67-80.

Moran, P.A.P. (1950), "Notes on Continuous Stochastic Phenomena," *Biometrika*, **37**, pp. 17-33.

Tipping, M. E., Bishop, C. M. (1999). Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology), **61**(3), pp. 611-622.

Woytinsky, W.S., Woytinsky, E.S. (1953). *World population and production. Trends and outlook*. The Twentieth Century Fund, New York City, (NY).