On the Use of Latent Variables in Education and Health Systems Evaluation

Giorgio Vittadini, Pietro Giorgio Lovaglio
Department of Quantitative Methods for Business Economic Sciences, Faculty of Statistics
University Bicocca-Milan, V.le Sarca 202, 20126 Milan , TEL +39 02 6448.3217, giorgio.vittadini@unimib.it

1. Introduction

The goal of transforming the EU into the most competitive and dynamic knowledge-based economy in the world within the year 2010 relies on the development of educational and training systems from both a worldwide perspective as well as from the perspective of lifelong learning.
In October 2001, the European Commission recommended the development of composite indicators for certain purposes within the Structural Indicators Exercise (European Commission 2001) which was reinstated by the report (European Commission 2002). The report identified and described 15 quality indicators in four areas: skills, competencies and attitudes; access to participation; resources for lifelong learning; and strategies and system development.
The European Commission (2007) reinitiated this inquiry more recently in a February 2007 communication. The *European Lifelong Learning Index* would become a composite indicator that could be used to track the state of the lifelong learning in the member states of the European Union.
To this end the main objective deals with the construction of synthetic (composite) indicators.
Composite indicators are valued for their ability to integrate large amounts of information into an easily understood format for a general audience. Despite their many deficiencies, composite indicators will continue to be developed, given their usefulness. They are highly appreciated in international comparisons, where it is often necessary to surmount national particularities and to consider to a common denominator.
Squeezing any complex system into a single metric, such as an index, creates a plethora of empirical challenges, such as data quality, indicator selection, and indicator importance.
When constructing a composite indicator of educational quality, the synthesis of simple indicators is an extremely sensitive procedure. In the community context in particular, the use of indicators aims at ranking statistical units, which can be negatively influenced by how the synthesis is constructed.
Furthermore, numerous problems affect the process of index construction, ranging from the selection of variables to the final result. The various dimensions of simple indicators creates the need for a normalization system which: respects their inherent nature (what they measure), their variability, and their weight (how direct their relation is with what they measure).
Moreover, there is partial overlapping among simple indicators, due to the correlation between them, so that it would be proper to extract their net contribution, rather than to treat them as a whole. This can be achieved by a pre-processing phase, aimed at selecting the most robust indicators, and by the choice of a suitable weighting system for the simple indicators.
Therefore, the selection of an appropriate methodology is a key procedural factor in summarizing the interaction of index indicators.
In recent years, statistical literature has addressed the challenges inherent to the construction of composite indicators. Basic methods of study design and data analysis are suggested by Cox et al. (1992), focusing in particular on the choice of measurement instruments and their properties, putting emphasis on weighting and aggregation. The authors suggest the need for long term observation periods and the presentation of changes over time, particularly for intervention assessment purposes geared toward revealing the natural variability of the composite indicator.
Toward this end, the Canadian Council of Learning has recently created the Composite Learning Index (CLI) developed to assess the state of lifelong learning over time for individual communities.
In fact, the central question addressed in the CLI report was the choice of a proper methodology for ensuring the strongest possible relationship between the composite indicators of learning and socio-economical outcomes.

The Composite Learning Index is a composite unobservable indicator, constructed as combination of four unobservable dimensions called pillars (Learning to Know, Learning to Do, Learning to Be, Learning to live Together, proposed by UNESCO and adopted by European Union policies concerning the Lifelong Learning program.), explored through its observable indicators which determine the socio-economical performance of a population.
In this perspective, CLI has been proposed within a statistical (causal) model framework, moving away from traditional weighting scheme methodologies since they do not take into account the causal links between the specified dimensions (e.g. Lifelong Learning indicators and socio-economic outcomes).
Thus, the primary question arising in this new perspective is linked to the choice of a proper methodological approach. One robust and realistic way to consider the construction of synthetic indicators, especially in educational and health systems evaluation, deals with the use and specification of latent variables (LVs), defined as unobservable dimensions whose manifestations are observed with errors through their observable indicators. The natural context involving causal models with non observable variables are the so called Structural Equation Models (SEM) with latent variables, or briefly, LISREL models (Everitt, 1994).
In next section, we furnish two main fields of application that methodologically justify the use of latent variables within a causal models framework.

2. The Use of Latent Variables in Educational and Health Systems Evaluation

The first application methodologically supporting the use of latent variables is specifically linked to Human Capital (HC). The re-launched Lisbon strategy places strong emphasis on knowledge, innovation, and the optimization of HC, which is closely linked to education, employment, and the health sector. The theory of human capital identifies human beings as the determining factor in income growth and a nation's wealth.
Over the last 50 years, the concept of HC or the human productivity factor has been systematically developed in economic literature. HC theory assumes that individuals will be paid in direct proportion their productive capabilities. This form of productivity typically depends on the level of education, educational achievement, work experience and on-the-job training, as well as job title, health history, parents' level of education, and socio-economical status. In this context, the pioneering efforts of Schultz, Mincer and Becker have been particularly noteworthy (see Becker, 1964), analyzing earning functions, also known as: age-earnings profiles, human capital production functions or Mincerian wage equations (Mwe). They specify that earned income for i-th worker depends on their level of education, usually years of schooling, plus a linear and quadratic term referring to their accumulated employment experience.
The known problems plaguing Mwe, such as the endogeneity of education and omitted individual ability, greatly depend on the relationship between HC individual stock and number of years of schooling; thus, the Mwe approach does not truly estimate human capital, because it only coincides with years of schooling: it would be unrealistic to conclude that two individuals having the same amount (years) of schooling also had the same amount of human capital.
Within a more pragmatic approach, HC can be reasonably considered as a broader multi-dimensional non-observable construct, depending on several and interrelated causes, indirectly measured by many observed indicators; briefly, it is assumed to be a latent variable (LV).
To this end, Dagum and Slottje (2000) have defined household Human Capital at a microeconomic level as the multidimensional non-observable construct generated by indicators as personal ability, home and social environments, investment in the education of the household head and spouse, whose effects are indirectly measurable by means of indicators, reflecting its effect, as earned income. They utilize Wold's PLS (Partial Least Squares, Wold 1982) to estimate HC as a linear combination of several observed indicators. Vittadini and Lovaglio (2007) propose obtaining HC as an unknown function of investment indicators and as a "latent cause" of performance indicators by

means of a system of structural equations, at disaggregated level (worker) by utilizing routinely available institutional educational and job market data flows.
The second example justifying the approach of latent variables arises in the managerial context. Kaplan and Norton (1992) proposed a model to support the decision-making process in business, called Balanced Scorecard (BSC) aimed at providing a multidimensional interpretation of business performance. The BSC is typically proposed to assess which areas (and to what extent) best contribute to the prospective of the creation of monetary value for statistical units (firms, companies, departments, etc). The conceptual scheme of Balanced Scorecard, using a balanced set of different types of indicators, has gained huge popularity, as it is an effective theoretical tool to support decisions regarding the multidimensionality of evaluation processes, the identification of individual and group targets, the allocation of resources, incentivizing, and the application of strategies concerning operational and organizational activities. The BSC, firstly, specifies the relations between the so called key performance areas (KPA) and secondly, specifies the observable indicators for these areas or Key performance index (KPI). Recently, in the health sector, the BSC has been applied in the statistical perspective (Lauro, 2007) to describe the mechanism producing creation of (monetary) value for health structures. In particular, it has been specified as a statistical model, hypothesizing that the economic performance (Economy KPA), inherent to the monetary-physical outputs of healthcare structures and their productivity depends on three macro-areas: Human Capital (referring to the quality of human resources in Health structures), Process (inherent to the quality and efficiency of healthcare processes) and Patient Satisfaction.
Since the four KPAs are unobservable multidimensional theoretical constructs measurable with errors by means of blocks of observed indicators (latent variables), the theoretical framework hypothesized by BSC can be naturally specified in a Structural Equation Models with latent variables.

3. Different Definitions of Latent variables.

Traditionally, the variables in a structural model (system of causal relationships) were directly observed variables, but recently several models have been studied which involve hypothetical constructs, i.e., latent variables (LVs), which, while not directly observed, have operational implications for relationships among observable variables. The observable variables or manifest variables (MVs) may appear as effects (indicators) of the LVs, as causes of the LVs, or as both effects and causes (Jöreskog and Goldberger, 1975; Jöreskog, 1981). In response to different situations, statistical literature has proposed two different definitions of LVs. The first definition is supplied by the LISREL Model, where there are “true LVs” in the sense of Bentler’s (1982) definition[1], sharing the Factor Analysis approach (Lohmöller, 1989),
In effect, CLI developers have used the LISREL framework to build a model characterized by strong relationships between the specified LVs, enabling an accurate approach. In particular, to adequately weight the individual pillars, the LISREL model (Joreskog, 1981) was applied in order to respect the latent nature of each pillar, to identify the measurement error for various observable indicators and to estimate causal relationships between the four pillars and the outcome indicators (Economic and Social Well-Being). In this way, the weighting of the indicators reproduced the observed variance-covariance matrix of manifest indicators (indicators for the pillars and outcome indicators).
Despite its widespread use, the LISREL model presents many drawbacks. Firstly, under general conditions, the LISREL model is not identified (Anderson and Rubin, 1956). Because of the complexity of the structural models it supports, large numbers of alternative but statistically equivalent, models can be supported by the same data: for example, reversing the direction of any

[1] A variable in a linear structural equation system is an LV if the equations cannot be manipulated so as to express the variable as a function of MVs only”. In this case the LVs are latent unobservable causes of MVs and “the dimensionality of the space spanned by the LVs is greater than the dimensionality of the space spanned by the MVs”.

causation path or replacing it with a correlation path will produce an equivalent model with the same fit indices (Stelzl, 1986).
Another problem in inferring a cause-effect in SEM is the complexity of its specifications: not specifying an important construct in the model and/or not specifying enough observed measurements for each construct (Bagozzi and Baumgartner, 1994) creates a bias resulting in an incorrect interpretation of the results, as in other types of statistical analysis (Hair et al., 1998). Finally, even if the parameters of the LISREL Model are perfectly identified, the scores of the LVs are not unique, i.e. there are infinite sets of latent scores for the same identified model (Bentler, 1982; Vittadini, 1989)
To resolve these drawbacks, a second definition of an LV arises in literature, defining it as an “Unobservable Component Variable" (Kmenta, 1991) of its observed indicators, estimated as linear combinations of its MVs.
In the context of causal relations between latent constructs, Partial Least Squares Path Modelling (PLSPM; Wold, 1982) is the counterpart of LISREL model embedding this second definition of LV. PLSPM proceeds in two stages: in the first stage, it estimates the latent scores as linear combinations of its own observed indicators (Wold, 1982), whereas in the second the parameters of the structural model (links between LV) and of the measurement model (links between MV and LV) are achieved through simple and multiple regressions, utilizing latent scores estimated in previous step. In this logic, the structural model becomes a model of Path Analysis between linear combinations of the observed variables. By iterating the steps of this procedure, the convergence of the algorithm presents the final estimation of the weights that estimates the scores of LVs as linear combinations of their observed indicators.
All of the techniques implemented in the PLSPM procedure work by extracting successive factors, or linear combinations of the predictors, that optimally address one or both of these two goals: explaining response variation and explaining predictor variation. In particular, principal components regression selects factors that explain as much predictor variation as possible, reduced rank regression selects factors that explain as much response variation as possible, while partial least squares regression balances the two objectives, seeking factors that explain both response and predictor variation (covariances between linear combinations of predictors and responses).
The LISREL and PLSPM approaches differ in the types of relationships they support between the observed variables and their associated latent constructs; in fact, there are two ways of relating the manifest variables to the respective latent variables: reflective and formative (Dijkstra, 1983). In the case of reflective indicators, typically specified in traditional factor analysis models, the latent constructs give rise to observed variables which covariate among themselves. In the case of formative indicators, latent constructs are combinations of observed variables and are not designed to account variance of observed variables. In that case, the model aims at minimizing residuals in structural relationships (explanation of unobserved variances).
In the European context, PLSPM has already been applied to both the European Economic Sentiment Indicator (Gelper and Croux, 2007), in order to gain insight into the beliefs of consumers, and to the European Customer Satisfaction Index (ECSI), a consumer satisfaction assessment instrument at both the macro (national) and micro (organizational) level (Tenenhaus et al., 2005; O'Loughlin and Coenders, 2004).

3.1 Discussion of the proper approach

The goal of LISREL (or hard modeling, due to its stringent hypothesis as normality of manifest variables and independence of errors) is actually to provide a statement of causality by seeking to find structurally or functionally invariant parameters (i.e. invariant features of the mechanism generating observable variables) that describe the mechanism of the process of data generation. Unfortunately, often real data does not meet these requirements. PLSPM (or soft modeling, avoiding the traditional assumptions for which the observations need to be independent and jointly sampled by the same distribution) works on real data.

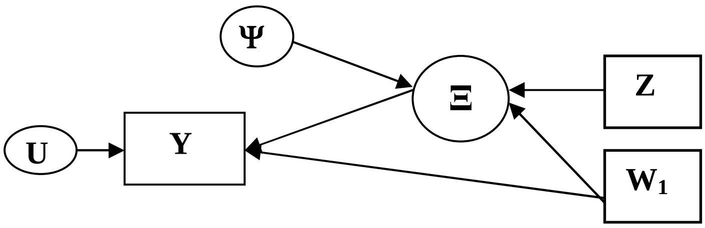
Since soft modeling creates optimal linear predictive relationships among variables, it aims at identifying predictive links, rather than causal ones. These relationships are interpreted as the best set of predictions available for a given study, considering all theoretical, measurement, distributional and practical limitations implicit in the data.
In some situations the LISREL approach is not appropriate yet the PLSPM approach is operational. The LISREL algorithm may also have limitations regarding to convergence, being better suited to research situations than to operational work. On the other hand, soft modeling is more robust. It can work with a few observations and numerous of variables, and it supports both types of observed variables (Tenenhaus, 1995), whereas the covariance-based SEM has been interpreted to support only reflective observed variables (Chin, 1998, Thompson et al., 1995).
Finally, soft modeling makes it possible to robustly estimate the latent variable scores of any model, obtaining consistent estimates and minimum variance predictions, without creating problems of either identification or indeterminacy.
Therefore, the focus of soft modeling is more on empirical data (measurement model) than on theory (structural model). As a matter of fact, the family of LISREL models is oriented to parameter estimation, thus yielding a better structural model because latent variables are space-free while soft modeling is oriented to the prediction of both manifest and latent variables.
For these reasons, in next paragraphs we will illustrate two different approaches for the achieving the estimation of structural models while avoiding the problems of the LISREL approach, particularly the non uniqueness of solutions. Section 4 illustrates the estimation of Human capital, simultaneously adopting a soft modelling approach to estimate latent scores, and a hard modelling approach to estimate causal relationships among LVs and MVs within a statistical model. Section 5 deals with the estimation of Balance Scorecard model applied in the Health sector by means of an advanced version of PLSPM.

4. Measurement model for the latent variable Human Capital

When considering HC as a latent variable (whose dimensionality requires investigation), we must first specify a measurement model which is consistent with a realistic economic process of HC accumulation, and then utilize a proper latent variable technique for the estimation of its scores. To this end, the schemes hypothesized in recent studies (Dagum and Slottje, 2000; Dagum et al., 2007; Vittadini and Lovaglio, 2007; Lovaglio, 2008) suggest that a set of indicators (e.g. containing earned income), called reflective indicators, is directly affected by the unidimensional latent variable HC, whose unobservable scores depend on specific indicators, called formative indicators (e.g. containing years of schooling), measuring the amount (money, time) of investment in HC.
Nevertheless, the question is then whether such specification, aiming to conceptualize the relationships among these blocks of observed variables, provides an accurate reflection of the full picture.
In fact, a more consistent system would take into account other exogenous socio-demographic factors (like sex, ethnicity, marital status, area of residence, and so on) which do not belong to the HC investment indicators set, but may have a causal impact on HC reflective indicators (endogenous). Statistically speaking, these factors are called concomitant indicators. Moreover, in many situations, as in the present case, these concomitant indicators may also have a causal impact on the scores of the LV specified in the model.
To take into account all considered factors, a consistent measurement model for HC must be invoked; in the general case of multidimensional HC, the path diagram assumes the specification of Figure 1. In this direction, the r-dimensional LV Human Capital, whose components are collected as columns of matrix $\mathbf{\Xi}$, admits two separated blocks of formative (or exogenous) indicators: specific indicators of investment in HC ($\mathbf{Z}$) and concomitant indicators ($\mathbf{W_1}$), which causally affect the HC scores and their reflective indicators as earned income ($\mathbf{Y}$).
Note that the specification of causal relations between concomitant indicators $\mathbf{W_1}$ onto HC investment indicators $\mathbf{\Xi}$ makes it possible to empirically explore the significance of the self-

selection process of education.

Figure 1 - Path diagram of adopted Structural model to estimate HC



Consistently with the path diagram of Figure 1, equations (1)-(2) specify the structural model and the measurement model for the r-dimensional latent variable HC, respectively:

$$\mathbf{Y} = \mathbf{W_1 C_1} + \mathbf{\Xi}\,\mathbf{\Lambda} + \mathbf{U} \qquad vec(\mathbf{U}) \sim (\mathbf{0}, \mathbf{I_n} \otimes \mathbf{\Theta}) \qquad (1)$$
$$\mathbf{\Xi} = \mathbf{Z G_1} + \mathbf{W_1 G_2} + \mathbf{\Psi} = \mathbf{W_2 G} + \mathbf{\Psi} \qquad \mathrm{rank}(\mathbf{W_2}) = r \qquad vec(\mathbf{\Psi}) \sim (\mathbf{0}, \mathbf{I_n} \otimes \mathbf{\Phi}) \qquad (2)$$

where $\mathbf{Y}$ is a (n,q) matrix of reflective variables, $\mathbf{W_1}$ (n,p) and $\mathbf{Z}$ (n,m) -embedded in a (n,m+p) matrix $\mathbf{W_2}$=($\mathbf{Z}$, $\mathbf{W_1}$)- matrices of formative observed variables, $\mathbf{C_1}$ a (p,q) matrix of regression parameters of $\mathbf{Y}$ onto $\mathbf{W_1}$, $\mathbf{\Lambda}$ a (r,q) matrix of regression coefficients between $\mathbf{Y}$ and the latent scores, collected in the (n,r) matrix Ξ=($\boldsymbol{\xi}_1$,...,$\boldsymbol{\xi}_r$), $\mathbf{G_1}$ (m,r), $\mathbf{G_2}$ (p,r) -embedded in a (m+p,r) matrix $\mathbf{G}$=($\mathbf{G_1}$, $\mathbf{G_2}$)- collecting weights of $\mathbf{Z}$ and $\mathbf{W_1}$ to define latent scores Ξ, $\mathbf{I_n}$ is an identity matrices of dimensions n, *vec* the operator staking columns, and $\otimes$ is the Kroneker product. In that specification, it is assumed that the dimensionality (r) of the latent variables Ξ, and thus that of the linear combination that defines it ($\mathbf{W_2 G}$), is less than m+p, the sum of $\mathbf{Z}$ rank and $\mathbf{W_1}$ rank. Finally, $vec(\mathbf{U})=(\mathbf{u_1},\ldots,\mathbf{u_i},\ldots,\mathbf{u_q})'$ is the random vector of errors in equations with zero means, zero covariances ($\mathbf{\Theta}$ is a q-dimensional diagonal matrix) and variances θ_i. Finally $\mathbf{\Psi}$ is the matrix of Ξ errors in variables, whose terms are typically correlated with unit variances and uncorrelated with vec($\mathbf{U}$) terms. Note that constraining r=1 and eliminating the $\mathbf{W_1}$ matrix, model (1)-(2) is equivalent to MIMIC model (Joreskog and Golberger, 1975).

4.1. Estimation of latent scores

The estimation of HC scores must be performed within complex causality relationships specified in the structural model. Figure 1 and model (1)-(2) demonstrates that latent scores must be estimated as combinations of HC investment variables $\mathbf{Z}$ (having causal impact onto reflective variables $\mathbf{Y}$), isolating two kinds of causal effects: the direct effect of $\mathbf{W_1}$ onto $\mathbf{Y}$ variables and the spurious effect of $\mathbf{W_1}$ onto $\mathbf{Y}$, by means of Ξ scores. Substituting equation (2) into equation (1) and aggregating terms, the reduced form of the model is:

$$\mathbf{Y} = \mathbf{W_1 C_1} + \mathbf{W_2 C_2} + \mathbf{V} \qquad \mathrm{rank}(\mathbf{W_2})=r, \quad \mathbf{G' W_2' W_2 G} = \mathbf{I_r} \quad vec(\mathbf{V}) \sim (\mathbf{0}, \mathbf{I_n} \otimes \mathbf{\Omega}) \qquad (3)$$

where $\mathbf{C_2}$=$\mathbf{G\Lambda}$ and $\mathbf{V}$=($\mathbf{\Psi\Lambda}$+$\mathbf{U}$). The reduced form suggests considering $\mathbf{W_2}$=($\mathbf{Z}$, $\mathbf{W_1}$) as a formative block of Ξ, but within $\mathbf{W_2}$, we need to remove the spurious effect of $\mathbf{W_1}$ onto $\mathbf{Y}$, following the rules of path analysis. Note that for identification purposes $\mathbf{W_2 G}$ is assumed to be orthonormal, and the errors' structures $\mathbf{\Theta}$ and $\mathbf{\Psi}$ (empirically indistinguishable) implied in equations (1)-(2) are absorbed in the full and not diagonal matrix $\mathbf{\Omega}$ (Lovaglio, 2008).

Latent scores in equation (3), (revealing the situation of an LV that admits a reflective ($\mathbf{Y}$) and two formative blocks, one full rank block ($\mathbf{W_1}$) and one deficient rank block ($\mathbf{W_2}$)) must be estimated after having removed the direct and spurious effect of $\mathbf{W_1}$ onto $\mathbf{Y}$. The proposed estimator belongs

to a set of models whose first aim is to explain a high proportion of response variation by linear components of predictor variables with structured errors, in presence of complex causal relationships.
The HC scores are estimated as that combination of $\mathbf{Z}$ and $\mathbf{W_1}$ indicators, net of the spurious contributions of $\mathbf{W_1}$ onto $\mathbf{Y}$ and net to the direct effect of $\mathbf{W_1}$ to $\mathbf{Y}$, essentially following Path Analysis rules, decomposing the correlation between observed variables in causal (direct, indirect) and spurious effects (for details see Lovaglio, 2008).

4.2 Application to HC model: the data

The proposed method has been applied to the estimation of HC for workers in the Milan area, based on administrative archives.
Available data refers to compulsory transmissions of subordinate contract employee declarations, mandatory for industries, and, recently, also for the Public Administration. The available database "*Employment Centers of the Province of Milan*" collects information about the longitudinal sequence of vocational experiences for workers with subordinate contracts in the private sector, who have registered a variation in their employment position between the years 2000 and 2005. In order to perform an HC analysis, this administrative archive is matched with administrative flows regarding income tax return declarations which provide information regarding annual gross earnings declared to official institutions. It collects workers' individual income tax returns, filed with the National Internal Revenue Service. This database covers the period 2000-2004, referring to workers' individual income tax return filed from 2001 to 2005.
We have considered 2004 as the year of reference (income declarations presented in 2005), because it is the most recent period regarding earned income. The utilized population refers to only 102,293 workers with residence in the City of Milan, because, at the moment, workers' individual income declarations are available only for workers with their declared residence in this area. The population is thus composed of workers, residents in the City of Milan, with subordinate contracts in the private sector, and vocational experiences recorded in the database of employment offices of the Province of Milan, having observed gross income in 2004.
As useful indicators (reflective, formative and concomitant) for HC analysis we have chosen the following: the unique HC outcome (reflective indicator) is 2004 gross earned income (workers' individual income tax return filed in 2005, from here on, *earned income*) composed of compensation of employees (including also transfers and economic assistance); as HC investment indicators we have considered: years of schooling (imputed by the last certification awarded), days of full-time work in the entire period of observation (2000-2004) and days of training in the period 2000-2003 (before 2004) and as concomitant indicators, only influencing earned income, we have considered: age, marital status, number of children, type of contract, industry, type of occupation (for the last three, the imputed values correspond to those having longest duration in 2004), whereas gender and nationality also have a causal impact on HC.
One main limitation of the available informative platform is the lack of information on worker careers prior to 2000 (and, in particular, the years of work experience), which we prefer to drop, instead of using the well known imputation rule, and to insert worker's age as proxy, as well as a lack of knowledge about parents' socio-economical status, while information about wealth of origin household is available only for a very limited number of workers living with parents, hence not considered.
Secondly, available training periods do not refer to training on the job, but to Lombardy regional courses (financed by European Social Fund, which typically target workers with low education groups) attended outside the workplace.
Looking at the composition of the population at 2004, 52% of workers were male. The age structure (mean age is 35) was homogeneous by gender, whereas the structure of contract types (homogenous by gender), 60% of workers have a permanent contract, 23% fixed term contract, 5% are workers in temporary work agencies (which constitutes the youngest worker group of mean age of 31 years

old), and 12%, other types of contracts. Considering the schooling level, 41% of workers (34% males, 46% females) have completed secondary school (8 years of schooling), 16% are technical high school graduates, 13% scientific or humanities high school graduates (16% males, 9% females), 7% vocational high school graduates (9% and 6%, respectively) and 15% are college graduates (same % for both sexes). Over 15,000 workers attended regional courses: the most-attended training typologies were post compulsory school training (27%), lifelong training (23%) and training courses to complete compulsory schooling (21%).

4.3 The results: the estimation of HC

With the methodology proposed in section 4.1 we obtain the estimate of standardized HC. The presence of qualitative variables does not represent an obstacle, because they can be expressed as dummy variables. The model's fit indicates that estimated HC and concomitant indicators explain 56% of the earned income. Table 1 displays the indicators of HC investment, the standardized regression coefficients (Std. coeff), their significance (Sig.) and, recalling that HC is estimated as exact linear combination of its formative indicators, the percentage impact of each indicator in the HC scores formation (%Weights).

Table 1- Coefficients and significance for HC indicators

HC Indicators	**Std coeff**	**Sig.**	**% Weights**
Years of Schooling	0.756	<.0001	57.1%
Days of Full-Time Work	0.564	<.0001	31.8%
Days of training	0.333	<.0001	11.1%

Note that the formation of HC is largely attributed to formal schooling, showing the largest impact (near 60%), followed by the stability component, measured by the period of full time work (32%), whereas the impact of the training duration is significant, but marginal.

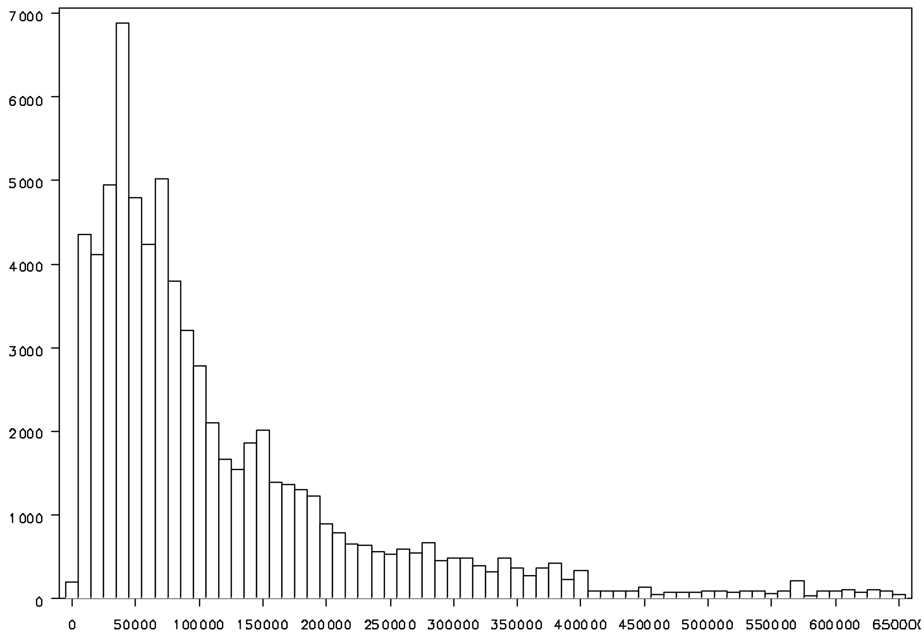
Table 2- Significance of HC and of concomitant indicators onto the labour income

Covariate	**F test**	**Sign.**
Human Capital	4641.54	<.0001
Gender	1232.88	<.0001
Age	1141.55	<.0001
Occupation	932.27	<.0001
Type of contract	420.06	<.0001
Number of Children	119.22	<.0001
Industry	73.93	<.0001
Marital Status	69.84	<.0001
Nationality	3.29	<.0001

Table 2 shows significant indicators affecting earned income: more than two thirds of explained income variance depends on factors that typically characterize the Italian labor market, such as gender, age and worker occupation, whereas less than one third is attributable to HC. The HC standardized coefficient shows an increase of earned income of 4,000€ for each additional HC standard deviation.

The final step of the analysis concerns the estimate of HC in monetary value; to this end we apply the actuarial method proposed by Dagum and Slottje (2000). In particular, starting by earnings' averages by workers' age, we construct the series representing the expected flow of earned income at each age (HC life cycle value), assuming a discount rate of 5% (approximately equal to Treasury bonds' interest) to actualize future earnings, capitalized by a rate of productivity (taking maximum value 3% at age 24, with a constant decrease in time until the age of 64, when it becomes null), and finally weighted by the survival probability to older ages.

Figure 2 – 2004 Human capital distribution of Milan workers



After having obtained the HC monetary mean μ (averaging the series of expected flows over ages), the estimated standardized LV distribution is exponentiated and translated in order to have mean μ. Figure 2 exhibits the monetary distribution of HC for the considered workers population, and Table 3 lists descriptive statistics for monetary HC and earned income distributions.
HC average is more than eight times higher than the average income, and HC distribution inequality is higher than income distribution inequality, confirming results in recent studies (Dagum and Slottje, 2000; Dagum et al., 2007; Vittadini and Lovaglio, 2007).

Table 3- HC and earned income descriptive statistics

Statistic	**Monetary HC**	**Earned income**
Median	79,757 €	13,907 €
Average	129,089 €	16,190 €
Gini ratio	0.501	0.435

Table 4 exhibiting correlations among estimated monetary HC and other relevant manifest variables, show reasonable values.

Table 4- Correlation matrix between HC, its indicators and earned income

	Monetary HC	**Earned income**	**Years of schooling**	**Days of full time work**
Earned income	0.450	1		
Years of schooling	0.602	0.278	1	
Days of full time work	0.453	0.264	0.092	1
Days of training	0.210	0.146	0.069	-0.157

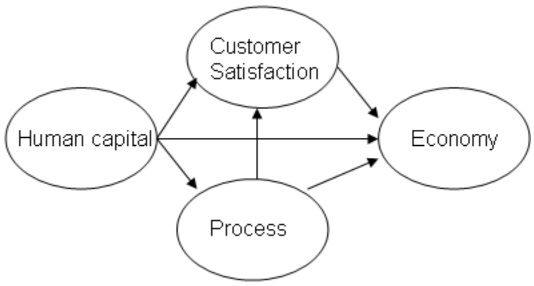
4.4 Discussion
The latent variable approach appears to be a consistent model, coherent with well-established economic theory which informs and supports the specification of relationships in the structural model.
Firstly, the estimation procedure effectively estimates both latent variables, respecting the economic definition (and the causal role in the structural model) of HC, as that stock of investment in

education and work experience (generated by formative indicators) to produce a sustained flow of income throughout the life span (reflective indicators). Secondly the LV scores have been estimated by removing the effect of the concomitant block (impact on earnings, but not on HC dimensions).
The empirical results of the Milan area have confirmed and highlighted the known situation regarding the *transition school to work* in the Italian context. Although HC results as being the most significant factor of earned income variability, it explains less than one third of explained labor income variance (56%), whereas an ample part is affected by dimensions linked to discrimination and career progression factors such as gender, age, and type of occupation, rather than education/training dimensions. These results are in accordance with knowledge regarding the Italian economy which is characterized by marked inequalities in the job market and weak incidence of education on the evolution of earned income, rather depending on automatic annual increments due to years of work experience and contractual seniority.

5. The BSC structural model

The BSC theoretical framework has been applied in the Lombardy region health sector for the assessment of the impact of health policies on business results of health structures within a pilot system (Lauro, 2007).
The scheme specified among latent dimensions is represented in the model shown in Figure 3.
The hypothesized diagram should be more complex, including other KPA such as Innovation or Finance, more articulated than Economy, or different perspectives of viewing the system (Stakeholders, physicians, regional authorities). However, the theoretical scheme has been proposed taking into account the available indicators in regional administrative archives.
As previously shown, the BSC from a statistical point of view is a structural model, estimable by PLSPM (Tenenhaus et al. 2005). Moreover, because the structure of observations in the health sector is typically nested (levels gradually more disaggregated of statistical units as Operative Units are nested in Presidium of the Corporate hospitals that are nested in Corporate hospitals nested in ASL), the latent variables scores have to be obtained taking into account the multilevel structure of the data to avoid an underestimation of the variability of the estimates and a distortion of fit statistics (Muthen, 1994), due to the correlation between units of the same group.

Figure 3 – Conceptual model of BSC for assessing corporate structures and nursing homes



This remark renders it possible to apply a statistical methodology called *Multilevel simultaneous component analysis* (MSCA; Timmerman, 2006) which obtains latent outcomes of BSC as Principal components (PC) of their observed indicators, allowing the separation of the overall variability (among all hospitals) in the separate contribution due to the groups (variability among ASL) and due to individual contribution (variability among hospitals in the same ASL). Secondly, the MSCA scores are integrated within PLSPM (PLSPM-MSCA), in order to estimate BSC model in a causal relationship framework.

5.1 Multilevel simultaneous component analysis (MSCA)

Let $\mathbf{Y_k}$ the (n_k,p_k) matrix of centered observed p_k indicators on n_k subjects for each LV, where k is the index of the m groups (k=1,..m), and N $(=\Sigma_k\, n_k)$ are the total number of observations.

Each $\mathbf{Y_k}$ block, describing the observations of p_k variables on the n_k hospitals belonging to the k-th ASL, are structured in vertical stacked blocks in the matrix $\mathbf{Y} = [\mathbf{Y_1},\ldots,\mathbf{Y_k},\ldots,\mathbf{Y_m}]'$.
Since the latent variables within the PLSPM are obtained as linear combinations of their observed indicators, the MSCA obtains, for each LV specified in BSC, two latent scores describing the variability within groups (hospitals net of ASL effect) and between different groups (ASL effect), in order to reproduce the maximum part of variability of the data. More specifically, MSCA decomposes $\mathbf{Y}$ in two parts:

$$\mathbf{Y} = \mathbf{T}_{(q)}\mathbf{A}_{(q)} + \mathbf{V}_{(v)}\mathbf{B}_{(v)} + \mathbf{E} \qquad (4)$$

where $\mathbf{T}_{(q)}$ describes the PCs within groups in a reduced space of dimension q, while $\mathbf{V}_{(v)}$ are the PC between groups in a reduced space of dimension v, $\mathbf{A}_{(q)}$ and $\mathbf{B}_{(v)}$ are loadings matrices and $\mathbf{E}$ is the matrix of residuals. Indices q and v are the chosen rank for the between and the within components less than the full rank dimensionality, $\Sigma_k p_k$ and m respectively. Equation (4) is a compact form specifying that each $\mathbf{Y_k}$ matrix can be decomposed in terms of $\mathbf{T}_{k(q)}\mathbf{A}_{(q)} + \mathbf{V}_{k(v)}\mathbf{B}_{k(v)}$ plus the residual term, where $\mathbf{T}_{k(q)}$,$\mathbf{V}_{k(v)}$ and $\mathbf{B}_{k(v)}$ are the k-th submatrices composing $\mathbf{T}_{(q)}$, $\mathbf{V}_{(v)}$ and $\mathbf{B}_{(v)}$, respectively.
MSCA has the fundamental property that the components $\mathbf{T}_{k(q)}$ and $\mathbf{V}_{k(v)}$, are orthogonal, within each block and for all groups.
In the MSCA the estimation of the principal components between and principal components within (and their respective loadings) are achieved by least squares criteria, minimizing the sum of squares (SSQ) of the following loss function (f_{MSCA}):

$$f_{\mathrm{MSCA}} = \boldsymbol{\Sigma}_k \, \mathrm{SSQ}\,[\mathbf{Y_k} - (\mathbf{T}_{k(q)}\mathbf{A}_{(q)} + \mathbf{V}_{k(v)}\mathbf{B}_{k(v)})] \qquad k=1,..m \qquad (5)$$

Considering that the members in parenthesis are orthogonal, their parameters can be obtained separately in each block. Thus MSCA finds the PCs and loading matrices essentially by separating each $\mathbf{Y_k}$ block in the following way:

$$\mathbf{Y_k} = \mathbf{Y_k} - n_k^{-1}\,\mathbf{I}_{nk}\mathbf{Y_k} + n_k^{-1}\,\mathbf{I}_{nk}\mathbf{Y_k} = [\mathbf{Y_k} - \mathbf{M}(\mathbf{Y_k})] + \mathbf{M}(\mathbf{Y_k}) = \mathbf{Y}^*_\mathbf{k} + \mathbf{M}(\mathbf{Y_k}) \qquad (6)$$

where $\mathbf{I}_{nk}$ is a square (n_k, n_k) matrix of unit elements, $\mathbf{M}(\mathbf{Y_k})$ is the matrix where each column contains the means of p_k variables for the k-th group and $\mathbf{Y}^*_\mathbf{k}$ is the matrix of differences between the observed values and the means of p_k variables for the k-th group.
With this decomposition, the PC between groups $\mathbf{V}_{(v)}$ are easily obtained by a simple Principal component Analysis (PCA) on the matrix $\mathbf{M(Y)}$ collecting all $\mathbf{M(Y_k)}$ matrices, whereas the latent scores within groups $\mathbf{T}_{(q)}$ by PCA on the matrix $\mathbf{Y}^*$ collecting all $\mathbf{Y_k}^*$ matrices[2].
In this way $\mathbf{T}_{(q)}\mathbf{A}_{(q)}$ describes the variability within the groups in a reduced space of dimension q while $\mathbf{V}_{(v)}\mathbf{B}_{(v)}$ describes the variability between groups in a reduced space of dimension v. The interesting property of MSCA is that the loadings matrix $\mathbf{A}_{(q)}$ in equation (6) is constrained to be the same for different k groups, otherwise the latent scores $\mathbf{T}_{(q)}$ obtained in different geometric subspaces would not be comparable among m groups.
Therefore, the obtained scores of the between structure express a synthetic measurement of the performance attributable to different hierarchical units of the second level (ASL), whereas the scores of the within structure express a synthetic measurement of the Hospitals' performance net to the clustering effect (ASL).

5.2 PLSPM- MSCA integration

[2] As measure of fit, the explained percentage of variance extracted by latent dimensions can be used. Since MSCA consists of independent sub models, three indices of fit can be defined: the percentage of variance explained by the within model, by the between model and by the whole model (Timmermann and Kiers, 2003).

The application of MSCA into the BSC model has the primary goal of decomposing the business performance variability of hospitals (estimation of the scores of the four areas) in the contribution due to the respective ASL and in its specific contribution, net of its inclusion in a particular ASL.
Nevertheless, the decomposition of overall variability between and within structures constitutes only the first step of the estimation of BSC model, because the estimation procedure has to be performed taking into account the causal relations between LVs in the structural model.
This is possible by applying the PLSPM algorithm separately to the within structure and to the between structure in order to obtain latent scores and causal parameters consistent with the supposed structural model separately in two different domains: that involving micro units (independently from belonging to a particular group) and that involving macro units (ASL).
The final product of a PLSPM model in a MSCA perspective in the BSC framework is the estimation of two structural models, one for the between and one for the within segment, revealing possible different causal structures, otherwise masked if the nested structure of observations were to be ignored.
More specifically, each data matrix of manifest indicators of the four KPA defining BSC, $\mathbf{Y}_{\mathbf{HC}}$ (Human Capital) $\mathbf{Y}_{\mathbf{PS}}$ (Patient satisfaction) $\mathbf{Y}_{\mathbf{P}}$ (Process) $\mathbf{Y}_{\mathbf{E}}$ (Economy) is reorganized in m vertically staked blocks, as $\mathbf{Y}_{\mathbf{HC}}=(\mathbf{Y}_{\mathbf{HCk}})$, $\mathbf{Y}_{\mathbf{PS}}=(\mathbf{Y}_{\mathbf{PSk}})$, $\mathbf{Y}_{\mathbf{P}}=(\mathbf{Y}_{\mathbf{P1k}})$, $\mathbf{Y}_{\mathbf{E}}=(\mathbf{Y}_{\mathbf{Ek}})$ and finally each block is decomposed in the between and within structure, as $\mathbf{Y}_{\mathbf{HCk}}=[(\mathbf{Y}^{*}_{\mathbf{HCk}}, \mathbf{M}(\mathbf{Y}_{\mathbf{HCk}})]$, $\mathbf{Y}_{\mathbf{PSk}} = [(\mathbf{Y}^{*}_{\mathbf{PSk}}, \mathbf{M}(\mathbf{Y}_{\mathbf{PSk}})]$,$\mathbf{Y}_{\mathbf{Pk}} = [(\mathbf{Y}^{*}_{\mathbf{Pk}}, \mathbf{M}(\mathbf{Y}_{\mathbf{Pk}})]$, $\mathbf{Y}_{\mathbf{Ek}}= [(\mathbf{Y}^{*}_{\mathbf{Ek}}, \mathbf{M}(\mathbf{Y}_{\mathbf{Ek}})]$.
Then we apply two separate PLSPM algorithms in the between structure on data matrix **B** collecting vertically m matrices $\mathbf{B}_{\mathbf{k}}=[\mathbf{M}(\mathbf{Y}_{\mathbf{HCk}}), \mathbf{M}(\mathbf{Y}_{\mathbf{PSk}}), \mathbf{M}(\mathbf{Y}_{\mathbf{Pk}}), \mathbf{M}(\mathbf{Y}_{\mathbf{Ek}})]'$ and in the within structure on the **W** matrix collecting vertically m matrices $\mathbf{W}_{\mathbf{k}}=(\mathbf{Y}^{*}_{\mathbf{HCk}},\mathbf{Y}^{*}_{\mathbf{PSk}},\mathbf{Y}^{*}_{\mathbf{Pk}}, \mathbf{Y}^{*}_{\mathbf{Ek}})'$.
Applying the PLSPM algorithm to the principal components obtained for two structures, the latent scores and parameters (causal links between LV) are not derived according to the MSCA criterion (maximization of the variability explained in each block) but by the maximization of regression criteria specified by PLSPM. In fact, the PC scores change depending on the causal relationships specified in the structural model, achieved as linear combination of adjacent LVs.
Thus, the application of the MSCA in each block of the four areas of BSC to the structure between and that within allows to obtain the scores of the four LVs (as principal components) which constitute the first step of the PLSPM algorithm (estimate of each LV as combination of its indicators) in a less arbitrary way with respect to what the same algorithm sets (arbitrary weights).
Then, the MSCA furnishes useful indications on the dimensionality of the LV in the two hierarchical structures, supporting evidence for possible different dimensionality between the two structures.

5.3 The PLSPM-MSCA for the Lombardy Health System

The data refers to a previous research entrusted to IReR (Lauro, 2007) by Lombardy Directorate of Health. This study, aiming at assessing the impact of health policies on business results in the Lombardy region was entrusted to the Regional Institute of Research (IReR) in 2007 (Lauro, 2007).
The research, involving 163 hospitals in the Lombardy Region, selected 24 indicators (Key performance index, KPI), representative of the four KPA in the BSC structure (Table 5).
Clearly, the chosen indicators do not cover the full picture of the Regional Health system, limiting its capacity for providing a realistic overall picture: for example Human Capital indicators concern only the training area, whereas the Process indicators do not cover information about safety, timeliness, accessibility, productivity and clinical efficiency of Health structures. Finally, indicators concerning the Economic area appear more consistent with qualitative rather than quantitative analysis.
Another drawback deals with the non availability of time series for their application in a longitudinal perspective. Nevertheless, the selected set of Table 5 constitutes the first attempt at an evaluation system for Italian Health sector.

Table 5- BSC - Key Performance Area and Key Performance Indicators

Area (KPA)	Sub-area	Indicator (KPI)	Label
Human Capital	*Training*	Internal training	RUFI
		External training	RUFE
Process	*Outcome*	Intra-hospital mortality rate	OMIN
		Mortality rate within 30 days	OM30
	Quality	Incidence of voluntary discharges	QDIV
		Impact of transfers	QTRS
		Impact of repeated hospitalizations	QRRP
	Health productivity and efficiency	Average extent of hospitalization (days) Days of hospitalization for doctor	EDGM EGDM
Pat. Satisfaction	*Organizational quality*	Hospitality	QOOS
		Staff	QOPE
		Hospital organization	QOOS
	Clinical Quality	Information	QCIC
		Hospital treatment	QCCU
		Discharge	QCID
	Satisfaction	Satisfaction versus Expectations	SATT
		Satisfaction on the health state	SSAL
Economy	*Economic result*	Income	ECMO
	Productivity	DRG income on days of hospitalization	PRNG
		DRG income upon number of cases	PRNC
		Budget income on number of employees	PRND
	Incidence of costs	Cost of work	ICLT
		Cost of health services	ISST
		Cost of hotel services	ISAT

To incorporate the nested structure of Lombardy Health, we have considered the presidium of the Corporate Hospitals (from here Hospitals) as the first level unit and ASL as the second level.
The elimination of the units containing one or more missing values has determined a final presence of 129 hospital structures (hospitals) in the operational database object of study.
In order to better compare the indicators, they have been normalized on a 0-100 scale (100 is the score for the best performance, whereas indicators concerning costs indicate a 100 vale for lowest cost).
The first model to be estimated is that referring to the *between structures* (ASL), in order to evaluate the performances of ASL (macro level units) in the BSC framework.
Before showing the estimated equations in the structural models by PLS in the between structure, we present results of the measurement models aiming at evaluating, for each LV, unidimensionality (all the MVs of a latent block must be image of a unique concept), mono-factorial validity (MVs of a block have to be higher correlated with their LV than other LVs), and discriminant validity (the part of variability that each LV shares with its block of indicators, called average communality AVE must be greater than the part of variability shared with the other LVs[3]).
Table 6 presents indices for controlling three properties. Apart from Economy (where ISST and ISAT have been eliminated), all blocks present highly satisfactory unidimensionality. Furthermore, mono-factorial validity and discriminant validity exhibit positive results, justifying the estimation of causal relations between LVs and their MVs.
Table 7 shows, for each endogenous LV, the path coefficients of exogenous LVs, the estimated bootstrap standard errors (Ds), the t statistic, its significance, the adjusted R squared (AdjR^2) and the contribution of each exogenous LV to the explained variance (%R^2).

[3] A block is considered unidimensional when the Cronbach's α (Cα) and Dillon-Goldstein's Rho (DG Rho) are greater then 0.7 and the discriminant validity for an LV holds if the squared correlations with other LVs are smaller than the average communalities, calculated on the indicators of its block.

Table 6- Unidimensionality, Mono-factorial validity and Discriminant validity (Between-structure)

Investigated Dimension	Index and indicators	Human Capital	Process	Cust. Satisfaction	Economy
Unidimensionality (Indices)	Cα	**0,932**	**0,742**	**0,980**	0.735
	DG Rho	**0.969**	**0.797**	**0.986**	0.710
Mono-factorial validity (correlations and AVE)	Human Capital	**1**	0.054	0.038	0.020
	Process	0.054	**1**	0.165	0.272
	Cust. Satisfaction	0.038	0.165	**1**	0.010
	Economy	0.020	0.272	0.010	**1**
	AVE	0.932	0.384	0.894	0.317
Discriminant validity (loadings)	RUFI	**0.986**	-0.212	0.224	0.175
	RUFE	**0.945**	-0.253	0.123	0.062
	SSAL	0.417	-0.450	**0.795**	-0.055
	QOOS	0.168	-0.259	**0.973**	0.178
	QOPE	0.201	-0.407	**0.974**	0.144
	QOOR	0.091	-0.292	**0.964**	0.142
	QCIC	0.116	-0.386	**0.982**	0.135
	QCCU	0.178	-0.496	**0.987**	0.082
	QCID	0.074	-0.387	**0.891**	-0.036
	SATT	0.224	-0.403	**0.981**	0.150
	EGDM	-0.215	**0.807**	-0.584	0.267
	OMIN	0.036	**0.592**	0.234	0.631
	OM30	-0.229	**0.682**	-0.366	0.357
	QDIV	-0.068	**0.415**	-0.209	0.004
	QTRS	-0.308	**0.832**	-0.406	0.315
	QRRP	0.016	**0.552**	-0.044	0.520
	EDGM	0.139	**0.230**	-0.103	0.196
	ECMO	0.145	0.478	0.172	**0.988**
	PRNC	-0.129	0.041	-0.020	**0.167**
	PRND	0.241	0.474	-0.437	**0.731**
	ICLT	-0.067	0.457	-0.191	**0.485**

We note that only the link between Customer Satisfaction and Human Capital does not result as being statistically significant. However, Adj R^2 associated to the first two equations demonstrate the weak predictive power of the exogenous LVs Processes and Customer Satisfaction. Therefore Process and Customer Satisfaction performances of Lombardy region ASL are poorly modeled in a Balance Scorecard framework. Instead, the third structural equation, explaining Economy area, shows acceptable results (Adj $R^2 = 0.43$); particularly significant in the contribution of Process area contributing to the 84% of the explained variance, followed by Customer Satisfaction (8%) and Human Capital (7%).

Table 7- Structural regression equation (Between-structure)

Endogenous LV	Exogenous LV	Path Coefficients	Ds	t-value	Pr > |t|	AdjR^2	%R^2
Process	Human Capital	-0.232	0.086	-2.692	0.008	0.05	100
Cust. Satisfaction	Human Capital	0.107	0.083	1.286	0.201	0.17	12
	Process	-0.381	0.083	-4.586	0.000		88
Economy	Human Capital	0.240	0.069	3.471	0.001	0.43	7
	Process	0.718	0.074	9.686	0.001		85
	Cust. Satisfaction	0.345	0.074	4.700	0.001		8

Table 8 shows, the goodness of fit index (GoF) associated to the between structure: the absolute GoF is a compromise measure between the goodness of fit of the measurement models (calculated as the mean AVE values over all LVs) and the goodness of fit of the structural model (calculated as the mean of R^2 over all endogenous model).

Therefore, given that only the equation concerning Economy area obtains significant results, the goodness of fit of the structural model (0.217) and of the global model (0.370) are very poor, despite the satisfactory goodness of fit of the measurement models (0.652).

Table 8- Goodness of fit index (Between-structure)

Type of Model	Index
Global	Absolute GoF = 0.370
Measurement model	Mean AVE = 0.652
Structural model	Mean R^2 = 0.217

In the second step we have estimated the within-group structure.
Table 9 shows that for all blocks the unidimensionality, the mono-factorial validity and the discriminant validity are respected (ISST and ISAT have been eliminated from Economy).

Table 9 -Unidimensionality, Mono-factorial validity and Discriminant validity (Within-structure)

Investigated Dimension	Index and indicators	Human Capital	Cust. Satisfaction	Process	Economy
Unidimensionality (indices)	Cα	0.805	0.902	0.880	0.725
	DG Rho	0.852	0.980	0.810	0.830
Mono-factorial validity (correlations and AVE)	Human Capital	**1**	0.097	0.002	0.051
	Process	0.097	**1**	0.040	0.216
	Cust. Satisfaction	0.002	0.040	**1**	0.150
	Economy	0.051	0.216	0.150	**1**
	AVE	0.746	0.399	0.861	0.562
Discriminant validity (loadings)	RUFI	**0.977**	0.051	-0.341	-0.256
	RUFE	**0.734**	0.024	-0.112	-0.051
	SSAL	0.085	**0.839**	0.218	0.252
	QOOS	0.071	**0.925**	0.121	0.310
	QOPE	0.066	**0.927**	0.110	0.292
	QOOR	0.015	**0.941**	0.162	0.349
	QCIC	0.097	**0.974**	0.162	0.393
	QCCU	0.033	**0.957**	0.169	0.347
	QCID	-0.034	**0.877**	0.287	0.462
	SATT	0.055	**0.975**	0.196	0.387
	EGDM	-0.281	0.207	**0.461**	0.345
	OMIN	-0.110	0.116	**0.666**	0.289
	OM30	-0.295	-0.009	**0.744**	0.269
	QDIV	-0.272	0.175	**0.780**	0.431
	QTRS	-0.161	0.100	**0.824**	0.284
	QRRP	-0.003	0.144	**0.325**	0.099
	EDGM	-0.070	0.298	**0.439**	0.124
	ECMO	-0.168	0.419	0.387	**0.912**
	PRNC	-0.144	0.154	0.113	**0.538**
	PRND	-0.136	0.049	0.168	**0.589**
	ICLT	-0.235	0.295	0.498	**0.883**

Finally, Table 10 and Table 11 show the results of the structural model of the within structure and goodness of fit indices, respectively. Also in this case only the third structural equation explaining the Economy area obtain significant results, but with modest fit ($AdjR^2$ =0.31); moreover, the contribution to the variance of Economy scores is more balanced among Process (52%) and Customer Satisfaction (39%), while in the first two equations, Human Capital is not statistically significant, contrary to the between structure model.

Table 10- Structural equation (Within-structure).

Endogenous LV	Exogenous LV	Path Coefficients	Ds	t-value	Pr > |t|	$AdjR^2$	$\%R^2$
Process	Human Capital	-0.311	0.081	-3.830	0.000	0.05	100
Cust. Satisfaction	Human Capital	0.048	0.053	0.899	0.185	0.10	11
	Process	0.199	0.086	2.318	0.011		89
Economy	Human Capital	-0.225	0.177	-1.265	0.104	0.31	9
	Process	0.465	0.147	3.155	0.001		52
	Cust. Satisfaction	0.387	0.128	3.026	0.001		39

Fit statistics confirm the weak performance of the global model, due to the modest contribution of the structural model.

Table 11- Goodness of fit index (Within-structure)

Type of Model	Index
Global	Absolute GoF = 0.314
Measurement model	Mean AVE = 0.642
Structural model	Mean R^2 = 0.153

5.4 Discussion

This application has shown that the BSC theoretical framework, specifying three causal relationships between four Key performance areas is too complex a model and not supported by enough empirical evidence to quantify the performance of the Lombardy Region Health System. This may be caused by the lack of useful indicators: in effect it is not surprising that Human Capital scores, collecting only two manifest indicators, do not bring a significant contribution to the causal relationships specified in the BSC structure.

However, the results concerning the measurement models and the Economy area structural equation suggest that a more simplified version of the theoretical model appears very promising, and would be even more so, in the presence of more complete data (in a longitudinal perspective and disaggregated at hospitals and ASL level).

We remark that the PLSPM methodology, proposed without taking into account the hierarchical context of the Health structures, has revealed no significant causal relationships among the four latent dimensions (Lauro, 2007). Instead, the PLSPM-MSCA model has shown that, utilizing proper statistical methodologies which take into account the hierarchical nature of the organization of Health Systems, it is possible to explore and estimate causal relationships among four KPA of the structural model consistent with the BSC framework.

To conclude, we remark that guidelines for a proper methodology of composite indicators also stress the usefulness of the sensitivity analysis and underline that an adequate allowance for uncertainty associated with summary measures should be indicated through the use of plausible ranges of rank for each institution. As shown, PLSPM provides a set of indices for checking the discriminant validity between different latent variables, the mono-factorial nature of manifest variables, testing the different aspects of the predictive relevance of the model with respect to the structural prediction (endogenous variables from exogenous ones), the communality (formative indicators from endogenous variables) and the redundancy (formative indicators from exogenous variables). Moreover, the assessment of the model in term of robustness and sensitivity can be obtained by methodologies such as cross-validation and specifically by jackknife or bootstrap replications.

6. The perspectives for EU archives

Typically, analyses concerning HC are performed only in a cross sectional framework (Census and survey data). However, many criticisms of the wage functions approach are typically imputed to the lack of information in a longitudinal perspective.

Although Mincer’s equations represent a pragmatic method of incorporating some of the major implications of the optimal HC models into a simple econometric framework, they can be applied to the limited information available in Census or cross sectional survey data (Willis, 1986, pp.542-543). The availability of longitudinal information of individuals’ earnings, associated with information about education, family background and personal characteristics could enable the estimation of statistical earnings’ functions and education rate of return in a more consistent way, but unfortunately such ideal data is seldom available.

To this end, since it is not possible to replicate data collection systems as complex and costly as a census every ten years, and survey panel data are very limited in the statistical panorama, administrative registers can be used as an alternative. In fact, while Census (survey samples) are complete (representative) only in the case of being drawn in a predetermined instant, administrative archives describe, by definition, all of the relevant units. Continuity, meaning the registration of an event at the exact moment in which it occurs, is a characteristic unique to administrative registers, where the Juridical Units are required to report every modification in real time. Moreover, the cost of collecting administrative data, recorded in computerized archives, is notably lower than one shot surveys, and also informative content, being mandatory, captures information with a low margin of uncertainty.

The limitations posed by statistical information drawn by updating identifying micro data from a census performed every ten years has stimulated various experimental efforts by national statistical institutions, aimed at integrating census data with that of administrative registers. In this perspective, the orientation of national statistical institutions and international regulations was aimed at distributing the utility of administrative data as widely as possible, while at the same time limiting the use of survey data which cannot be described indirectly by administrative registers, and is inspired by the criteria of being all inclusive, manifesting continuity, cost consciousness, and the responsibility of the respondents (Statistics Canada, 1988; Eurostat, 1997, 1998).

Therefore, countries such as Denmark, France, Switzerland and Canada utilized administrative registers as the basis for reorganizing their statistical systems for industries, (Statistics Canada, 1988; Eurostat, 1997), during the nineties.

In 1998, the Official Italian Institute of Statistics was already integrating identifying micro data from the Statistical Archive of Active Firms (ASIA) with the partial data from analyses originating from the main statistical surveys conducted on industries, creating an informative statistical system which provided a significant improvement in data quality, enriching and broadening analysis opportunities. Above all, the primary advantage of these experiments, conducted on five large archives and in 31 Italian Provinces, has been the progressive implementation of complex statistical procedures, such as normalization, linkage, optimum modality choice, estimation of missing data and control of data quality (Martini, 2000).

In the Health Sector, the objective of developing national evaluation systems has stimulated various experimental efforts by national statistical institutions.

Although international Institutions as WHO and OECD suggest some main goals of these evaluation systems (microeconomic efficiency, macroeconomic sustainability, social equity, accessibility to health services), there is, nevertheless, ample heterogeneity among national systems, reflecting particular situations and different priorities.

Table 12 illustrates some of the most significant examples of evaluation systems proposed in national health sectors.

The various countries appear strongly heterogeneous especially due to the quality and amount of available data on different dimensions to be investigated. However, instead of building perfect systems of indicators, it appears to be more constructive to define an approach aiming to specify some common goals and proper indicators able to measure them in a rigorous statistical way.

In this perspective, a first attempt was made within the European Community Health Indicators (ECHI, 2005) project, an international research carried out in the framework of the Health Monitoring Programme and the Community Public Health Programme 2003-2008.

In the ECHI project, an initial set of 40 indicators used by WHO-Europe, OECD and Eurostat in their international databases was proposed in the period 1998-2004, after their arrangement following the ECHI taxonomy. Initially a “short list”[4], ECHI (2005) proposed a longer list after

[4] Note that only 8 are potentially useful in our analysis: 33 Hospital beds; 34 Physicians employed; 35 Nurses employed; 37 Hospital in-patient discharges, limited diagnoses; 38 Hospital in-patient discharges, limited diagnoses; 40 Surgeries: PTCA, hip, cataract; 41Expenditures on health; 42Survival rates breast, cervical cancer.

February 2004, containing a few items for which regular and comparable data collection is still possible within a very short period[5].

Table 12 - European evaluation systems in Health sector

Country	Project	Main Objectives and Characteristics of the Project
France	COMPAQH	Health system improvement, Health process quality control, (variability reduction of Process Results) Evaluation of Health performances, comparison of selected hospitals performances Process and Outcome Indicators.
Germany	Hospital Improvement	Health process quality control , Comparison between selected hospitals performances, Indicators on health financing
Holland	Sneller Better	Transparency, Effectiveness and Quality of Care improvement: Knowledge and accountability increase, benchmark, improvement and good practices diffusion.
Spain	Quality Evaluation	Health sector responsibility: adverse events and patient safety, Insurance information, transparency and reporting.
Switzerland	Health European Research Area project	Adverse events analysis, reporting and cause analysis
United Kingdom	Star Rating and Performance Indicators	NHS health system description and health structures improvement, Health structure effectiveness, efficiency and customer satisfaction, Health structures ranking and benchmark (Star system).

Nevertheless, the main limitation of this data deals with the disaggregation level. In fact, proper evaluation models able to control selection bias and adverse selection in the health sector, such as Multilevel models (Aitkin and Longford, 1986; Goldstein and Spiegelhalter, 1996), utilized mostly in the framework of risk-adjusted evaluations, invokes microdata that have to be available at a fine

[5] Among them we present indicators and areas useful for our goals. They are ***- Human Capital:*** *Health Care Resources and Manpower***:** Emergency services availability; Employment in general health administration; Employment in provision and administration of public health programmes; Employment in retail sale and other providers of medical goods; Health services employment; Hospital staff ratio: acute care; Hospitals employment; Midwives employed; Nurses employed; Nurses staff ratio: acute care; Physicians by specialty. *Education***:** Dentists graduated; Nurses/midwives graduated, Pharmacists graduated. ***- Process:*** *Outcomes:* 30-day mortality rate following AMI; 30-day mortality rate following CABG; 30-day mortality rate following stroke; Antibiotic resistance; Avoidable deaths; Cancer survival rates: breast, cervix, colorectal; Coverage of cancer registration; Decubitus in nursing and elderly homes; Iatrogenic disease/death; Major amputations in diabetics; Renal failure in diabetics; Surgical wound infection; *Quality and Health care process indicators***:** 28-day emergency re-admission rate; Access of care for children; Compliance with good oncology practice; Delay of cancer treatment; Diabetes monitoring; Emergency services response time; Emergency services: advanced interventions; Equity of access; Femur fractures waiting time; Health promotion in hospitals; Parental accompaniment in hospitals; Quality of blood products; Retinal exams in diabetics; Stage at cancer diagnosis; Support to women in the prenatal period; Very preterm births outside NICU; Waiting times. *Clinical productivity and Health care utilisation*
In-patient care utilisation: Average length of stay, limited diagnoses; Bed days acute care; Hospital admissions asthma; Hospital in-patient discharges, limited diagnoses; Occupancy rate, acute care. *Out-patient care utilisation*: Day case-discharge ratio, limited diagnoses; Emergency services by diagnosis; Emergency services high priority; Emergency services utilisation; General practitioner utilisation; Hospital day cases, limited diagnoses; Medical specialist contacts; Occupational safety and health services use. ***- Patient Satisfaction:*** *Health care quality/performance***:** Responsiveness of the health system; Satisfaction of mothers with prenatal care; Satisfaction with the health care system.
- ***Economy:*** *Health expenditure and financing***:** Expenditure on personal health care; Expenditures on collective health care; Total/public/ private expenditures on health**.** *Expenditure on medical services:* Expenditure on home care; Expenditure on in-patient care; Expenditure on out-patient care; Expenditures on ancillary services**.**
*Medical goods dispensed to outpatients***:** Expenditure on pharmaceutical goods and other medical non-durables; Expenditures on medical appliances/other durables. *Total health expenditure by age group:* Expenditure by age group; Expenditure by disease group; Expenditures on occupational health and safety. *Health expenditure by fund source***:** Expenditures by fund source; financial equity/access indicator**.**

grained level and, following the authors’ experience, at least at hospital level (outcome and structural characteristics).

7. Conclusions

Recalling the question contained in the Canadian Learning Index report “What methodology will ensure that the composite of a particular set of indicators has the strongest possible relationship with the broader outcomes?” it appears fundamental to create the specification of proper statistical models involving non-observable and multidimensional constructs, especially in the education and health fields. The two applications presented in this study have been focused on realizing this new perspective.

The availability of official statistics at the EU level or administrative data bases makes it possible to operatively utilize these models for building evaluation composite indicators on particular aspects of educational and health systems.

In the HC application, we have proposed a consistent technique for the estimation of the latent variable HC, specified in a realistic measurement, by utilizing routinely administrative archives. The methodology has properly estimated HC as the rank-one best linear combination of their formative indicators (investment indicators) that best fits their reflective indicators (accounting the effects of investment), net to the direct effects of concomitant indicators (indicators affecting income but not contained in the block of HC investment indicators) to reflective indicators. Nevertheless, available administrative archives must collect additional information about workers’ educational performance, as measures of quality of education as well as characteristics of schools or University attended and on training on the job.

The second application has illustrated how a solid conceptual scheme such as the Balance Scorecard can be proposed to assess the causal relationships between Human Capital, Process production, Patient Satisfaction and Economy results in health sectors.

The analysis has shown that the PLSPM model in an MSCA perspective allows the exploration and estimation of two structural models, one for the between and one for the within part, revealing different causal structures, otherwise masked if the nested structure of observations were be ignored.

The synthesis obtained by PLSPM-MSCA procedure can be extended in many different situations or structures. For example, longitudinal data, showing a clear hierarchy structure, may be used to estimate the between component as a static component (hospitals mean performances on the considered period) and the within component as a dynamic component (hospitals trends in temporal instants for each hospital).

References

Aitkin M., Longford N. (1986) Statistical modelling issues in school effectiveness studies, *Journal of the Royal statistical society*, Series A, 149, 1-43.

Anderson T.W., Rubin, H. (1956) Statistical Inference in Factor Analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5: J. Neyman, Ed. Berkeley University of California Press.

Bagozzi R.P., Baumgartner H. (1994) The Evaluation of Structural Equation Models and Hypothesis Testing, in R. P. Bagozzi (ed.) *Principles of Marketing Research*, Blackwell, Cambridge MA, 386-422.

Becker G.S. (1964) *Human Capital* . New York : Columbia University Press.

Bentler P.M. (1982) Linear System with Multiple Levels and Types of Latent Variables, in K. Joreskog and H. Wold, (eds.), *System under Indirect Observation: Causality, Structure, Prediction*, North Holland, Amsterdam, 101-130.

Chin W.W. (1998) The Partial Least Squares Approach to Structural Equation Modeling, in G.A. Marcoulides (ed.) *Modern Methods for Business Research*, Erlbaum, NY, 295-336.

Cox D.R., Fitzpatrick R., Fletcher A.E., Gore S.M., Spiegelhalter D.J., Jones D.R. (1992) Quality of life Assessment: Can we keep it simple? *Journal of the Royal Statistical Society*, Series A, 155, 353-393.

Dagum C., Slottje D.J. (2000) A new method to estimate the level and distribution of household Human Capital with application. *Structural Change and Economic Dynamics*, 11, 67-94.

Dagum C., Vittadini G., Lovaglio P.G. (2007) Formative indicators and effects of a causal model for household Human Capital with application. *Econometric Reviews*, 26, 579-596.

Davis P., Tso M. (1982) Procedures for reduced rank regression. *Applied Statistics*, 31, 244-255.

Dijkstra T. (1983) Some comments on maximum likelihood and Partial Least Squares Methods, *Journal of Econometrics*, 22, 67-90.

European Commission (2001) Education council report to the European council on the "concrete future objectives of education and training systems" (doc. 5980/01).

European Commission (2002) Detailed work programme on the follow-up of the objectives of education and training systems in Europe (OJC 142 of 14.6.2002).

European Commission (2007), *A coherent framework of indicators and benchmarks for monitoring progress towards the Lisbon objectives in education and training,* Communication from the Commission.

ECHI (2005), *International Compendium of Health Indicators*, European Community.

EUROSTAT (1997) Proceedings of the Seminar on the Use of Administrative Sources for Statistical Purposes: Luxembourg, 15-16 January 1997.

EUROSTAT (1998) Use of administrative sources for business statistics purposes, *Handbook of good Practices*, Version 0, Luxembourg.

Everitt B.S. (1984) An introduction to latent variable models, Chapman and Hall, N.Y.

Gelper S., Croux C. (2007) The predictive power of the Euro-pean Economic Sentiment Indicator (ESI), *FETEW Research Report*, KBI_0707, K.U.Leuven.

Goldstein H., Spiegelhalter D.J. (1996) League tables and their limitations: statistical issues in comparison of istitutional performance, *Journal of the Royal statistical society*, 159, 385-443

Hair J.F., Anderson R.E., Tatham R.L., Black W.C. (1998) *Multivariate Data Analysis with Readings,* 5th Edition. Englewood Cliffs, NJ: Prentice Hall.

Jöreskog K., Wold H. (1982) *System under Indirect Observation*, North Holland, Amsterdam.

Joreskog K. Goldberger S. (1975) Estimation of a model with multiple indicators and multiple causes of a single latent variable, *Journal of American Statistical Society,* 70, 631-639.

Kaplan R.S., Norton D.P. (1992) The balanced scorecard. The measures that drive performance, *Harvard Business Review*, January-February, 171-179

Kmenta J. (1991) Latent Variables in Econometrics, *Statistica Neederlandica*, 45, 73-84.

Lauro N.C. (2007) Valutazione dell'impatto delle politiche. Realizzazione di un sistema pilota in ambito sanitario, IReR, 2005A023, Milano.

Lohmöller J.B. (1989) *Latent Variables Path Modelling with Partial Least Squares*, Physica-Verlag, Heidelberg.

Lovaglio P.G. (2008) Process of Accumulation of Italian Human Capital, *Structural Change and Economic Dynamics*, 19, 342-356.

Martini M. (2000) L'integrazione statistica di registri e indagini. In C. Filippucci (Ed.), *Tecnologie informatiche e fonti amministrative nella produzione di dati*, Franco Angeli, Milano, 291-305.

Muthén B. (1994) Multilevel covariance structure analysis, *Sociological Methods & Research*, 22, 376-398.

O'Loughlin C., Coenders G. (2004) Estimation of the European Customer Satisfaction Index, Maximum Likelihood versus Partial Least Squares. Application to Postal Services, *Total Quality Management & Business Excellence*, 15, 1231-1255.

STATISTICS CANADA (1988) Proceedings of "Statistical uses of Administrative data: an international Symposium", 23-25 November 1987, Statistics Canada press.

Stelzl I. (1986) Changing a Causal Hypothesis without Changing the Fit: Some Rules for Generating Equivalent Path Models, *Multivariate Behavioral Research,* 21, 309-331.

Tenenhaus M., Vinzi V., Chatelin Y.M., Lauro C. (2005) Partial Least Squares, *Computational Statistics & Data Analysis*, 48(1), 159-205.

Tenenhaus M. (1995) *La Régression PLS: Théorie et Pratique*, Editions Technip, Paris

Timmerman M.E, Kiers H.A.L. (2003) Four simultaneous component models of multivariate time series from more than one subject to model intraindividual and interindividual differences, *Psychometrika*, 86, 105-122.

Timmerman M.E. (2006) Multilevel component analysis, *British Journal of Mathematical and Statistical Psychology*, 59, 301-320.

Thompson R., Barclay D.W., Higgins C.A. (1995) The Partial Least Squares Approach to Causal Modeling: Personal Computer Adoption and Use as an Illustration, *Technology Studies: Special Issue on Research Methodology*, 2, 284-324.

Van der Leeden R (1990) *Reduced Rank Regression with Structured Residuals*. DSWO Press, Leiden.

Vittadini G. (1989) Indeterminacy Problems in the LISREL Model, *Multivariate Behavioral Research*, 24 (4), 397-414.

Vittadini G., Lovaglio, P.G. (2007) Evaluation of the Dagum-Slottje method to estimate household Human Capital. *Structural Change and Economic Dynamics*, 18, 270-278.

Willis R. (1986) Wage determinants: A survey and reinterpretation of Human Capital earnings functions, in O. Ashenfelter and R. Layard (eds.), *Handbook of Labor Economics*, Elsevier Science Publishers, 525-602.

Wold H. (1982) Soft modelling: the basic design and some extensions, in K. Jöreskog and H. Wold, H. (eds.) *Systems under indirect observation: causality, structure, prediction*, North Holland, 1-54.