IRISS at CEPS/INSTEAD

An Integrated Research Infrastructure in the Socio-Economic Sciences

HUMAN CAPITAL ESTIMATION THROUGH STRUCTURAL EQUATION MODELS WITH SOME CATEGORICAL OBSERVED VARIABLES

by

Annamaria Di Bartolo

IRISS WORKING PAPER SERIES

No. 2000-02

IRISS-C/I

An Integrated Research Infrastructure in the Socio-Economic Sciences at CEPS/Instead, Luxembourg

A project supported by the European Commission, the Ministry of Culture, Higher Education and Research (Luxembourg) and the National Science Foundation (USA)

RESEARCH GRANTS

for individual or collaborative research projects
(grants awarded for periods of 4-6 weeks)

Who may apply?

IRISS-C/I supports collaborative and/or internationally comparative research in economics and other social sciences. We encourage applications from all interested individuals who want to carry out research in the fields of expertise of CEPS/Instead. A separate funding agreement is in place for US researchers. Their access to IRISS-C/I is funded by the National Science Foundation.

What is offered by IRISS-C/I?

Free access to the IRISS-C/I research infrastructure (office, computer, library...);
Access to the CEPS/Instead archive of micro-data;
Technical and scientific assistance;
Free accommodation and a contribution towards travel and subsistence costs.

Research areas

Survey and panel data methodology; income and poverty dynamics; persistent and new poverty; gender, ethnic and social inequality; unemployment; segmentation of labour markets; new forms of work/a-typical employment; education and training; social protection and redistributive policies; impact of ageing populations; intergenerational relations; effects of family-related policies; regional development and structural change.

Additional information and application forms

IRISS-C/I
B.P. 48 L-4501 Differdange (Luxembourg)
IRISS-C/I homepage: http://www.ceps.lu/iriss/iriss.htm

Human Capital Estimation through Structural Equation Models with some Categorical Observed Variables.

Annamaria Di Bartolo•
Dipartimento di Statistica, Università degli studi di Milano-Bicocca,
Via Bicocca degli Arcimboldi, 8. 20126 Milano. anna.dibartolo@unimib.it

Abstract: The aim of this paper is to estimate, for US, Canada and Italy, the latent variable human capital and its causal relationship with labor income, through some Structural Equation Models. The analyzed models contain some observed categorical variables, which imply the use of the two-stage estimation technique.

Key Words: Human Capital, Structural Equation Model (SEM), Polychoric Correlation, Weighted Least Squares, LISREL.

1. Introduction: Human Capital Definition.

The Department of Economic Affairs of the United Nations (1953) defined *investment in human capital* as investments made to increase the productivity of the labor factor.
A country's future production can be developed, not only by increasing the conventional capital stocks, but also through investments in education and on-the-job training, immigration, acquisition of knowledge, and improvement of health and life standards of the workers as well as many other intangible factors that affect the productivity of labor.
Since both conventional capital and human capital involve costs and promises of future earnings, it is possible to recognize a symmetry between the two concepts. *Human capital is a estimation of the ability of a person to produce labor income.*
Therefore, human capital support policies should include: 1) promotion of educational projects and scholarships, 2) development of research, 3) improvement of social and family life standards, and 4) development of political tools to control immigration.

Human Capital estimates (individual or aggregate) have been applied in economics for the determination of the dynamics of the employment earnings market (earning functions)[1], for the analysis of the income distribution, for the investigation of the economic growth and for the measure of the social costs of emigration. But, despite the wide definition given, in most empirical studies human capital is estimated solely by the education level of the subject. The reliance on education level is due, in part, to the difficulties of measuring to components of human capital.
Therefore, researchers trying to estimate human capital face two problems that are not easy to resolve. The first is a definition problem. The second is an evaluation problem.
As already noted, some authors use years of schooling as an estimate of human capital. Although they may want to use a wider definition, some researchers have to deal with the deficiency of information available from surveys.

• This paper has been presented at the "*International Workshop on Correlated Data: estimating function approach*", Trieste, Italy (22-23 October '99). In November '99, the author received a grant from the European Commission, TMR program, Access to Large Scale Facilities and will be hosted by IRISS-C/I at CEPS-INSTEAD, Differdange (Luxembourg) to continue the present research.
[1] Mincer (1970), Di Bartolo(1999a).

The aim of this paper is to study the personal productivity of human capital for US, Canada, and Italy. The concept of human capital will coincide with the productivity of the human factor. The productivity considered depends on the abilities, the education, the level of satisfaction, and the opportunities that society and family have provided to the individual. It should not depend on social-demographic factors, like: sex, ethnicity or marital status.

Excluding the possibility that only one variable adequately describes human capital, it is natural to define the human capital or, more formally, the productivity of the human factor, as a latent variable.

The productivity variable can be measured rather effectively by various indicators. For example: level of education, educational achievement, work experience and on-the-job training, job title, health history, parents' level of education, and economic class of origin could predict productivity.

Considering the human capital a latent variable implies the use of a latent variable technique for the analysis. In this job I analyze the relation between human capital and labor income using structural equation modeling which is a novel approach[2].

Classical studies on earnings affirm that, among the factors that influence individuals' labor income, there are variables as sex, ethnicity, civil status[3]. Those variables should not have impact on human capital, defining it as a collection of abilities and results of investments in personal productivity. If a relationship between these demographic factors and human capital is detected, their influence could be attributed to the social structure of the analyzed countries.

Figure 1 describes these hypothesized relationships. The scheme outlines the hypothesis that personal income is directly imputable to a latent variable, called "Human Capital Productivity", which depends on the individual's characteristics and ability and a latent variable, called "Social and Demographic Factors". Between the two latent variables there is a further relation that needs to be investigated[4].

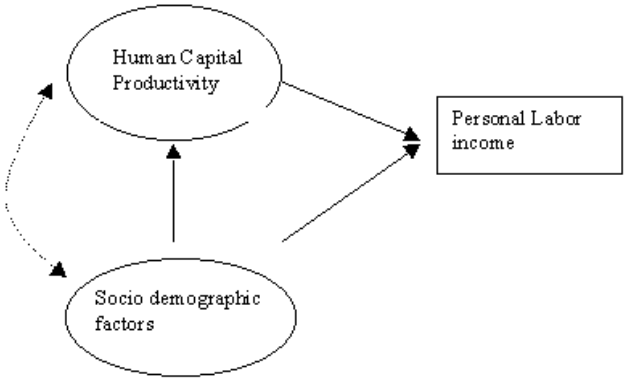

Figure 1 Relationship among human capital, labor income and other socio-demographic factors

This paper presents a preliminary investigation of the causal relationships described in Figure 1. I formalized and estimated human capital models for US, Canada and Italy through structural equation models. The data, collected from official investigations of personal and families' income[5], were supplied by the Luxembourg Income Study (LIS database).

Although some of the analyzed observed variables are categorical, the latent variables are all continuous. The choice of the variables included in the model has been limited (and forced) by the lack of more adequate indicators. The parameter estimation was computed in LISREL 8.3 utilizing, the two step procedure described in the following paragraph.

---

[2] Dagum (1994) estimated US human capital latent variable using Partial Least Squares.

[3] Becker (1993), Mastrodonato (1991)

[4] It could be a covariation or a dependency relationship.

[5] Italy '95: Indagine Campionaria sui Bilanci delle Famiglie, Banca Italia; US '94: March Current Population Survey, FBS; Canada '94: Survey of Consumer Finances.

2. Structural Equation Models with Categorical Observed Variables: Definition, Problems and Solutions.

Most researchers in applied statistics think in terms of modeling the individual observations. Structural equation modeling (SEM) "procedures emphasize covariances rather than cases" (Bollen 1989, p.1).

Structural equation modeling is a multivariate technique combining aspects of multiple regression (examining dependence relationships) and factor analysis (representing unmeasured concepts or factors with multiple variables) to simultaneously estimate a series of interrelated dependence relationships.

Five steps characterize most applications of SEM: (1) model specifications[6], (2) identification[7], (3) estimation, (4) testing fit and (5) respecification.

The continuous latent variable structural equation model can be expressed in two parts: a measurement model and a structural (or latent) model.

The measurement model consists of a multivariate regression model that specifies how the latent variables depend upon or are indicated by a set of observed variables (indicators or measures). It thus describes the measurement properties (reliabilities and validities) of the measures.

The structural model specifies the causal relationships among the latent variables and between latent variables and independent observed variables (or background variables). It also describes the causal effects, and partitions the variance into explained and unexplained groups. The links between variables are summarized in the model parameters: loadings, structural coefficients and variance-covariance structure of errors and latent variables. The parameters can be classified as free, fixed or constrained.

The fundamental hypothesis for the SEM methods is that the covariance[8] matrix of observed variables is a function of a set of parameters. If the model was correct and if we knew the parameters, the population covariance matrix would be exactly reproduced. This fundamental relation can be expressed by the following equation

[1] $S=S(q)$

where S, is the population covariance matrix of the p observed variables included in a specific model, $S(q)$ is the covariance matrix implied by the model and q is a vector containing the free parameters of the model.

In SEM, the unconstrained parameters of a proposed model are estimated by minimizing a discrepancy function (F) between a consistent estimate **S** of the unknown population covariance matrix and the covariance matrix implied by the model ($\mathbf{C}=S(q)$).

Browne (1984) demonstrated that the best known discrepancy functions (Maximum Likelihood (ML), Generalized Least Squares (GLS) and Unweighted Least Squares (ULS)[9]) are all, asymptotically, special cases of a generic discrepancy function:

[2] $F = (\mathbf{s}-\mathbf{c})' \mathbf{W}^{-1} (\mathbf{s}-\mathbf{c})$

where **s** and **c** are vectors of $u = p(p+1)/2$ elements obtained by placing respectively the nonduplicate elements of **S** and **C** in a vector, $\mathbf{W}^{-1}$ is a positive definite matrix of order u x u

[6] This step is usually depends on one's theory or past research experience. Anderson and Gerbing (1988) suggest a two step approach for a correct model formulation, based on a confirmatory factor analysis of the observed variables.

[7] Identification determines whether it is possible to find unique values for the parameters of the specifies model. Ridgon E. (1995).

[8] Correlation matrixes can be analyzed instead of covariance matrixes. Cudek (1989).

[9] Bollen (1989) pp. 104-115 and pp. 333-335.

obtained by inverting a positive definite weight matrix **W.** To estimate the model parameters q**,** the weighted least squares discrepancy function [2] is minimized with respect q**.** Under very general assumptions[10], if the model holds in the population and if the sample covariance matrix **S** is a consistent estimate of S, any specification of the [2] obtained with a positive definite **W** will give a consistent estimator of q**.**

Further assumptions must be made, however, if one needs an asymptotically correct chi square measure of goodness-of-fit and asymptotically correct standard errors of parameter estimates. Browne (1982) demonstrated that if **W** is chosen to equal or to be a consistent estimate of the asymptotic covariance matrix of **s** with **s**, then the parameters estimator obtained is asymptotically efficient within the class of functions that fall under [2]. In this case **W** is defined the correct weight matrix. Browne thus suggested an "asymptotically distribution free" (ADF) discrepancy function, where the correct weight matrix **W** is based on direct estimation of the fourth-order moments of the residuals[11]. If the observed variables have a multivariate normal distribution, or if **S** has a Wishart distribution, the GLS and ML functions are special cases of equation [2] and lead to asymptotically efficient parameter estimates utilizing a less demanding computational procedure.

When the distributional assumption is false, however, these discrepancy functions are, in effect, operating with incorrect weight matrices.

If the model was specified correctly and the distributional assumptions for the data were satisfied, analysts could use a test statistic with an asymptotic chi-square distribution to test null hypothesis that the specific model leads to an exact reproduction of the population covariance matrix of the observed variables. A significant test statistic would cast doubt on the model specification. Jöreskog (1967) sounded an early warning about overinterpretating the chi-square statistic. The use of chi square as a central chi-square statistics, in fact, is based on the assumption that the model holds *exactly* in the population. This may be an unreasonable assumption in most empirical research. A consequence of this assumption is that models, which hold *approximately* in the population, will be rejected in large samples[12]. In some cases it would be more reasonable to assume that the model holds approximately and then try to assess the error of approximation in the population as proposed by Browne and Cudeck (1993).

The described theory has been developed for structural equation model with continuous observed variables. Jöreskog (1993) affirms "ordinal variables are not continuous variables and should not be treated as if they are…they require other techniques than those that are traditionally employed with continuous variables[13]."

Several covariance structure analyses of ordinal variables reported in literature are based either on product moment correlations or on polychoric correlations and the maximum likelihood estimation method. In both cases, this can lead to incorrect results and invalid conclusions. When one or more observed variables are ordinal, the product moment sample correlation matrix **S** is not a consistent estimator of S**.** With ordinal variables, estimates of the polychoric correlation should be computed rather than Pearson's correlations. However, if a SEM is estimated by ML method applied to polychoric correlations, the parameter estimates are consistent, but standard errors of parameter estimates and chi-square goodness-of-fit measures are asymptotically incorrect.

Theory and application of structural equation model when some or all of the observed dependent variables are ordinal have been considered by several authors, for example Muthén (1984), Lee, Poon and Bentler (1990) and Jöreskog (1993, 1994).

---

[10] Browne (1984), Bentler (1983) and Bollen (1989) p.428.

[11] Browne's development is a theory for sample covariance matrices for continuos variables. In practice, correlation matrices are often analyzed. Several authors, for example Jöreskog and Sörbon (1996) have extended Browne work deriving the correct weight matrix form in case of correlation matrices and non continuos indicators.

[12] Even trivial deviations of a model from the actual structure could be detected and could lead to a rejection of the null hypothesis.

[13] Bollen (1989) pp.433-446 examines the model assumptions violated with categorical indicators and the consequence of violations.

The estimation of the model should be done in two steps.
The *first step* involves estimating polychoric, polyserial, tetrachoric correlations[14] for the observed variables or rather for the underlying continuos response variables. In fact it is necessary to assume that there is a continuos unobserved variable *x** underlying the observed ordinal variable *x*. It is the underlying variable *x** we were interested in, not the observed variable *x*. The hypothesized relation between *x* and *x** is

$$[3]\quad \Pr\{x = i\} = \Pr\left\{ t_{i-1} < x^* \le t_i \right\}, \quad i = 1,2,...,k$$

where *k* is the number of categories and $t_0 = -\infty$, $t_1 < t_2 < ... < t_{k-1}$, $t_k = +\infty$ are theresolds that have be estimated defining a probability distribution function for *x**. The origin and the unit of measurement in *x** is arbitrary. For mathematical convenience it is common to choose $x^* \sim N(0,1)$ even if other distributional assumptions could be made[15].
The *second step* estimates the parameters of the model by weighted least squares [2] using a correct weight matrix which must be a consistent estimate of the asymptotic covariance matrix of the polychoric, polyserial, tetrachoric or Pearson correlations estimated in the first step. Different formulas for the weight matrix have been given by Muthèn (1984), by Lee , Poon and Bentler (1990) and Jöreskog (1993) and respectively included in MPLUS, EQS and LISREL.

Unlike traditional normal theory methods, the described *two steps approach* can yield unbiased, consistent and efficient parameter estimates. It however presents some problems[16].
The first step assumptions are theoretically reasonable only in some cases. Even if for many attitude items, a researcher could be interested in the relationship among the normally distributed, continuos underlying response variables, for other continuously distributed variables, such as "current drug use" ("yes" vs. "no"), it is difficult conceive of a normally distributed underlying response variable. Moreover, some variables, as gender, are inherently categorical, so no continuos underlying variable could exist.
The second step estimation method, although ADF, has also some significant limitations. The estimation of the correct weight matrix places severe practical limits on the number of variables that can be considered (maximum is about 25). The use of the WLS estimator also requires that large samples be used (at least 500-1000 cases, depending on the complexity of the model).

In testing structural models with some ordinal observed variables Muthén (1994) suggests to consider two levels of testing. The first level of testing concerns tests of the distributional assumptions for the continuous variables that are hypothesized to underlie the categorical variables. Jöreskog (1999) indicates some possible remedies when the underlying bivariate normality does not hold at the first step of the described estimation method, here summarized: (a) assume that polychoric correlations are robust against violations of underlying bivariate normality, (b) dichotomize all variables and use tetrachoric correlations, (c) assume the model is only approximate and use Browne and Cudeck (1993) ideas to asses the degrees of approximation.
The second level of testing involves the structural equation model's validity and corresponds most closely to the conventional testing for continuos variables. It can be therefore verified using the classical tests statistics or the ones suggested by Browne and Cudeck (1993).

---

[14] Polychoric correlations between ordinal variables, tetrachoric correlations between dichotomous variables , polyserial correlations between ordinal (or dichotomous) and continuous variables and Pearson's correlations between continuous variables must be estimated.
[15] Jöreskog (1993) p.166
[16] West, Finch and Curran (1995) pp.68-70

3. Italy, US and Canada Human Capital Models: Some Results.

Utilizing Luxembourg Income Study datasets, structural equation models of human capital for US, Canada and Italy have been specified following five steps listed in the previous paragraph and the model building strategy suggested by Anderson and Gerbing (1988).

The utilized datasets present the following sample sizes: US 1994: N = 73533, Canada 1994: N = 46065 and Italy 1995: N = 5640. The analyzed units are *worker in the labor force*. Eight observed variables has been selected to investigate the relationships described in figure 1: years of schooling (*yrs*), age (*age*), marital status (*ms*), living place (*lp*), working position (*pos*), sex (*sex*), ethnicity (*eth*) (not available for Italy), logarithm of labor income (*ly*). A description of those variables is given in Table 1. *yrs* and *pos* have been chosen as human capital latent variable (*hcf*) indicators. The lack of additional meaningful human capital indicators represents a limit of this analysis. Pursuing the specification rules of LISREL the income observed variable (*ly*) is assumed to be a perfect measure of the income latent variable (*income*^{o}*ly*) The other observed variables are included in the models as background variables influencing both human capital and labor income.

For every country, the initial model, including all the variables, has been modified and tested again using the same data[17]. The goal was to find a model, which not only fits the data from a statistical point of view but which also has a meaningful interpretation for every parameter.

The first level of testing allows the acceptance of the polychoric correlations estimates assuming that those estimates are robust against violations of underlying bivariate normality.

Table 2 shows exact fit and close fit statistics for the final models.

Table 1 Variables Description.

Variable	*Name*	*Note:*
Human capital productivity	*hcf*	Latent continuos variable
Years of schooling	*yrs*	Years of schooling necessary to complete the level of education.
Working position	*pos*	0-3 depending on the level of responsibility required.
Age	*age*	From 15 to 65 years
Marital status	*ms*	Married=1, unmarried=0
Sex	*sex*	Male=1, female=0
Ethnicity	*eth*	Caucasic=1, other=1
Living place	*lp*	Small urban or rural center=0, big urban center =1
Logarithm of labor income	*ly*	Total labor income before taxes for US and Canada, after taxes for Italy.
Income	*Income*	Latent continuos variable coinciding with *ly*

Table 2. Exact fit and close fit statistics.

	US '94	**CANADA '94**	**ITALY '95**
Sample Size	73533	46065	5640
Minimum Fit Function Chi-Square	2376.29	250.521	58.746
Model Degree of Freedom	9	6	5
Chi-Square for Independence Model	238293.704	159281.314	28456.473
Independence Model Degrees of Freedom	28	28	15
Root Mean Square Error of Approximation (RMSEA)	0.0598	0.0297	0.0437
90 Percent Confidence Interval for RMSEA	(0.0578 ; 0.0618)	(0.0267 ; 0.0329)	(0.0341 ; 0.0540)
P-Value for Test of Close Fit (RMSEA < 0.05)	1.000	1.000	0.837

[17] The model generating approach has been applied (Jöreskog 1999).

The exact fit hypothesis should be rejected for every model. However, it is important to note that in large samples the power of the test can lead to a rejection of the null hypothesis when it is true. Moreover the distributional assumptions do not hold in the analyzed data. The test of close fit (RAMSEA) allows the acceptance of all the three models. For all chosen models parameter estimates, standard errors, squared multiple correlations, coefficients of determination and correlation of estimates are reasonable, proving a close fit of the model to the data.

Table 3 shows the estimate results for the measurement model of the latent variable human capital.

Table 4 presents the parameters estimates of the structural model. Socio-demographic variables' influence on human capital productivity and income has been investigated through some background variables. In the model selection process, some causal relationships have been deleted because these relationships were not significant. These insignificant relationships vary by country which may indicate dissimilar socio-economic structures in the countries considered.

Table 3. Estimation results of Human Capital Measurement Model

	US '94			**CANADA '94**			**ITALY '95**		
	Loading	***Error var.***	***R²***	***Loading***	***Error var.***	***R²***	***Loading***	***Error var.***	***R²***
yrs	0.505	0.745	0.255	0.511	0.739	0.261	0.820	0.327	0.673
Std. Err.	0.00887	0.00794		0.0165	0.0173		0.0244	0.0502	
t-value	56.893	93.876		31.023	42.683		33.552	6.514	
pos	0.716	0.487	0.513	0.865	0.251	0.749	0.956	0.0867	0.913
Std. Err.	0.0113	0.00849		0.0245	0.033		0.0256	0.0415	
t-value	63.653	57.372		35.362	7.623		37.265	2.09	

Table 4. Estimation results of Structural Model

		hcf	***age***	***lp***	***ms***	***sex***	***eth***	***Error var.***	***R²***
US '94	***hcf=***		+0.445*age	+ 0.134*lp		- 0.594*sex		0.444	0.556
	Std. Err.		0.0122	0.00441		0.0137			
	t-value		36.498	30.426		-43.456			
	income=	+0.398*hcf			+ 0.071*ms	+ 0.446*sex	+ 0.063*eth	0.829	0.171
	Std. Err.	0.0134			0.0028	0.0111	0.0038	0.00747	
	t-value	29.736			25.801	40.063	19.649	110.876	
CANADA '94	***hcf=***			+0.179*lp	+0.154*ms	-0.333*sex	0.148*eth	0.826	0.174
	Std. Err.			0.0217	0.0161	0.0251	0.0203		
	t-value			8.261	9.576	-13.247	7.265		
	income=	+0.296*hcf	+0.103*age	+0.121*lp	+0.202*ms	+0.373*sex	+0.0476*eth	0.76	0.240
	Std. Err.	0.0227	0.00820	0.0160	0.0132	0.0203	0.0148	0.0176	
	t-value	13.051	12.514	7.58	15.237	18.372	3.225	43.126	
ITALY '95	***hcf=***		+0.191*age	+0.183*lp,	+0.0967*ms			0.881	0.119
	Std. Err.		0.0395	0.0252	0.0438				
	t-value		4.845	7.249	2.205				
	income=	+0.378*hcf			+0.275*ms			0.732	0.268
	Std. Err.	0.0178			0.0112			0.0237	
	t-value	21.259			24.629			30.851	

Age is often interpreted as a proxy of experience. It could explain why there is a causal relation among *age* and *hcf* in US and Italy models. However it is necessary remember that often this interpretation could be incorrect, especially for individuals that delayed their entry to the labor market.

The effect of *lp* on *hcf*, could indicate that in urban centers there are more opportunities for investments in human capital productivity.

In Italy's model *sex* effect on *income* and *hcf* is not significant[18]. In US and Canada's models it appears in both the structural equations. For both the countries it is interesting to note that *sex* (male=1, female=0) has negative structural coefficient in the human capital equation.

An indicative *estimate* of the human capital variable can be obtained through *regression factor scores*[19]. The regression factor scores are the OLS estimates of the regression coefficients from the "hypothetical" regression of the latent variable on the observed variables.
The results of the regression factor scores obtained through the specified final models are presented in the following three equations:

[4] *US '94:* *hcf= 0.177*yrs+ 0.383*pos+0.125*ly+0.261*age-0.009*ms+0.079*lp-0.008*eth-0.404*sex;*

[5] *Canada '94:* *hcf= 0.148*yrs+ 0.739*pos+0.084*ly- 0.009*age+0.023*ms+0.036*lp+0.034*eth-0.118*sex;*

[6] *Italy '95:* *hcf= 0.180*yrs+ 0.792*pos+0.016*age+0.037*ly-0.002*ms+0.015*lp;*

For US data, the final human capital model, the distribution of the human capital estimates and the years of schooling variable distribution are respectively presented in figure 2, 3 and 4.

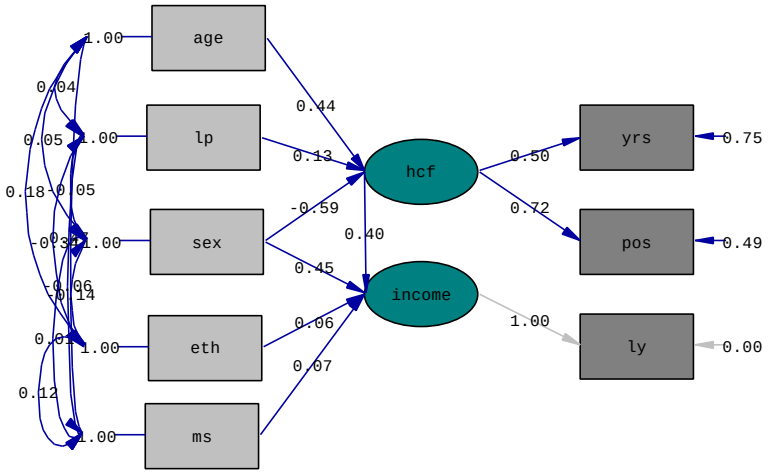

Figure 2 US '94 Human Capital Model Path diagram

[18] In the model building process the variable *sex* was eliminated because it was not significant.
[19] Bollen (1989) p.305.

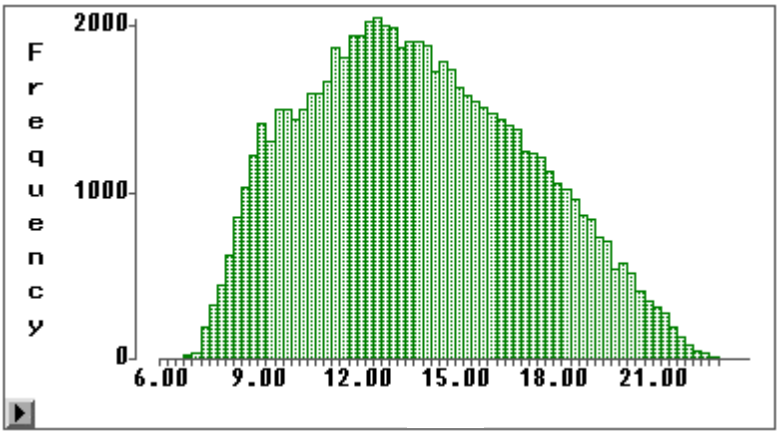

Figure 3 US '94: Human Capital Estimates Distribution.

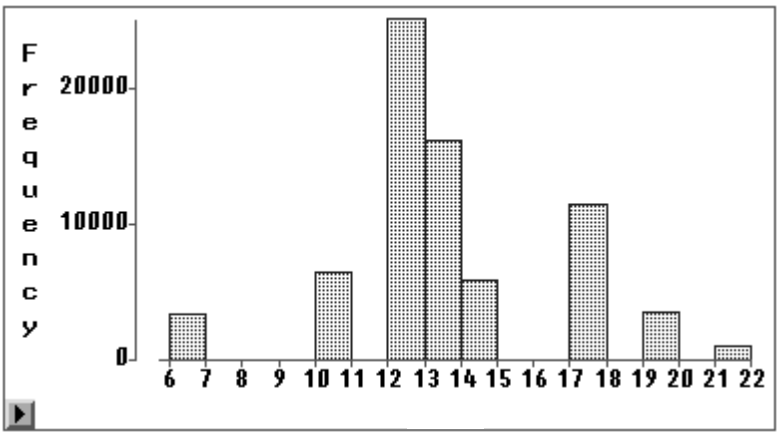

Figure 4 US '94: Years of Schooling Distribution

Even if *yrs* is a very good indicator of human capital productivity, the comparison between figure 3 and 4 provides evidence that it should not be used as only measure of the latent variable human capital.

4. Conclusions.

Even though the analyzed models include just a small number of the significant indicators the preliminary results are very interesting. The model structure in Figure 1 seems to fit to the data (better for USA and Canada than for Italy). The effects of the socio-demographic variables on income are in some way stronger than the human capital one, for the three countries. Although this analysis has produced a preliminary estimate of the personal human capital distribution, the subject needs further investigation for the following reasons:

1) The number of indicators needs to be expanded. More variables will more accurately describe the relations investigated.
2) Some utilized indicators are categorical. Structural equation models with categorical variables require the use of specific estimations methods. The theory is still under development (in particular fitting analysis techniques) and it has never been applied to human capital models.
3) It would be extremely interesting to extend the analysis to other European countries. The comparison results will generate an improved understanding of the labor market dynamics across countries.

5. References.

Anderson, J.C. & Gerbing, D.W. (1988). Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach, *Psychological Bulletin*, 103(3), 411-423

Becker, Gary S. (1993) *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Chicago: The University of Chicago Press, Third Edition.

Bentler, P.M. (1983) Moment Structure Models, *Journal of Econometrics*, 22, 13-42.

Bollen, K.(1989) *Structural equations with Latent Variables*. Wiley, New York.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 1-21.

Browne, M.W., Cudek, R.. (1993) Alternative Ways of Assesing Model Fit, in K. Bollen, e J.S Long. (eds.) *Testing Structural Equation Models*, SAGE Pubblications, 205-234.

Cudek, R. (1989) Analysis of correlation matrices using covariance structure models, *Psychological Bulletin*, 2, 317-327.

Dagum, C., (1994) Human Capital, income and wealth distribution models and their application to the U.S.A., in *Proceedings of the Business and Economic Statistic Section of the American Statistical Association*, 154th Meeting, 253-258 .

Di Bartolo, A. (1999) *Definizione e Metodi di Stima del Capitale Umano*, Tesi di Dottorato di Ricerca, XI ciclo, Università degli Studi di Bologna.

Di Bartolo, A. (1999a) Modern Human Capital Analysis: Estimation of USA, Canada and Italy Earning Functions, *Working Paper*, Dipartimento di Statistica, Università di Milano-Bicocca.

Jöreskog, K.G. (1967) Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32(4), 443-482.

Jöreskog, K.G. (1993) Latent Variable Modeling with Ordinal Variables, in K Haagen, D.J. Bartholomew & M. Deistler (eds.), *Statistical Modeling and Latent Variables*, Elasevier Science Publisher, 163-171.

Jöreskog, K.G. (1994) On the estimation of polychoric correlations and their asymptotic covariance matrix, *Psychometrika*, 59(3), 381-389.

Jöreskog, K.G. (1999) Aspects of Structural Equation Models, *Lecture Notes of Structural Equation Modeling with LISREL Class*, IES Spring 1999, Professional Development Training Session, Glacher Center, University of Chicago, Chicago.

Jöreskog K., Sörbon D. (1996) *Lisrel 8: User's Reference Guide*, Scientific Software International, Chicago.

Lee, S., Poo W. & Bentler P.M. (1992) Structural equations models with continuos and polytomous variables, *Psychometrika*, 57(1), 89-105.

Mastrodonato, Antonio (1991) *I Capitali Umani*, CEDAM, Padova.

Mincer, Jacob (1970) The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach, *Journal of Economic Literature*, 1-26.

Muthén, B. (1984) A general structural equations model with dichotomous, ordered categorical, and continuos latent variable indicators, *Psychometrika*, 49, 115-132.

Muthén, B. (1993) Goodness of fit with categorical and other nonnormal variables, in K. Bollen, e J.S Long. (eds.) *Testing Structural Equation Models*, SAGE Pubblications, 205-234.

Rigdon, E.E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, 30(3), 359-383.

Trivellato U. (1990) Modelli di comportamento e problemi di misura nelle scienze sociali: alcune riflessioni, in: SIS, Atti della XXXV Riunione Scientifica, Padova 18-21/4/1990, Cedam, 1, 11-34.

United Nations, Department of Economic Affairs, (1953) "Concept and Definitions of Capital Formation", Studies in Methods, series F, No. 3.

West, S.G, Finch, J.F. & Curran, P.J. (1995) Structural Equation Models With Nonnormal Variables, in R. Hoyle (ed) *Structural Equations Modeling: Concepts Issues and Applications*, Sage Publications, 56-75.