

Session Number: First Poster Session
Time: Monday, August 23, PM

*Paper Prepared for the 31st General Conference of
The International Association for Research in Income and Wealth*

St. Gallen, Switzerland, August 22-28, 2010

Measuring Human Capital with Latent Variables

Barbara Liberda and Anna Grochowska

For additional information please contact:

Name: Barbara Liberda
Affiliation: Warsaw Univeristy

Email Address: Barbara.Liberda@uw.edu.pl

This paper is posted on the following website: <http://www.iariw.org>

I. Introduction

Human capital is undoubtedly a key element of today's economies. A new phase of society has begun, transforming social structures into information societies relying on information and knowledge as the key inputs into nearly all economic processes. Creation, diffusion and manipulation of information information and expertise becomes a significant economic, political, and cultural activity. Wealth, power and freedom of action are derived from access to and effective use of knowledge. The markers of this new form of knowledge society are not only technological and economic but also occupational, spatial and cultural.

II. Human capital – definition and measurement

It is widely agreed in the human capital literature (Becker, 1974; Mincer, 1958) that investments in human capital, such as formal education, on-the-job training or migration raise observed earnings at older ages, because returns are part of earnings then, and lower them at younger ages, because costs are deducted from earnings at that time. Some persons earn more than others because they invest more in themselves. The general theory of human capital has a wide variety of important applications: it explains phenomena such as interpersonal and inter-area differences in earnings, the relation between age and earnings and the benefits of specialization (Becker, 1974).

Despite the important role of human capital in modern societies, there are still many unknowns about the process of its accumulation and its micro- and macroeconomic determinants. Human capital is often proxied as attained education or monetarily - in terms of individual earnings. In such a quantitative form, human capital is a component of individual welfare, production or growth functions.

Expressing human capital in value terms is sufficient to measure and compare it across individuals, however, does not explain its causes. In this study we focus on incentives to acquire skills. Contrary to common approaches, which concentrate on

individual and global benefits from human capital, we focus exclusively on its determinants.

Educational attainment is affected by a variety of micro- and macroeconomic factors. Government structural policies, social policies, credit constraints, taxation, education subsidies and other institutional incentives have a significant impact on human capital accumulation (Fleischhauer, 2007). Given the specific character of the dataset used (socio-economic survey data of selected Polish regions), this article concentrates on microeconomic factors shaping human capital attainment, such as family background, environmental surroundings, access to physical, communication infrastructure and aspirations.

"Human capital" can be defined as knowledge, skills, attitudes, aptitudes, and other acquired traits contributing to production (Goode, 1959). Education can occur through, but is not limited to, formal educational activities such as preschool programs, schools, and formal training programs (Behrman, 2009). The proximate determinants of education are inputs into knowledge and skills. According to some authors (Blundell, Dearden, Meghir & Sianesi, 1999) there are two main components of human capital with strong complementarity: early ability (whether acquired or innate) and skills acquired through formal education or training on the job. Inborn abilities and acquired skills are two manifest elements of human capital that we will explore in this paper.

In line with the literature on this subject, we decided to add one more (third) component of human capital – health. "Nothing in the concept of human capital implies that monetary incentives need to be more important than cultural and non-monetary ones", wrote Becker in 1974. Many studies document the positive impact of education on health (eg. lower probability of smoking), civic attitudes (rising propensity to voting) and even on the sensitivity to art and love for nature.

Many disciplines put forward the relation between biological and behavioral linkages such as chronic disease or disability with their implications for individual and community welfare. Micro- and macroeconomic researchers are striving to understand the connections between economic development and the demographic

transition. Numerous studies document the causal relationships between economic development and health outcomes. The relationship between conditions under which people live and their expected life span and health status refer to “health production functions.” The relationship between an individual's stock of health and well-being can be called “health human capital return” (Schultz, 2010).

In the framework of a dynamic forward-looking model of human capital investments, education experiences can be determined by a series of family and individual decisions given past, current and expected future resources, markets and policies. To understand the nature of inputs and incentives related to education, attention must be paid to both the demand and the supply of education investments, which depend on the decisions of education suppliers, the preferences of education consumers, situation in the labour markets (on-the-job training) as well as public policy choices (Behrman, 2009).

However, not only the education component of human capital but also individual health endowments can be explained using the forward-looking model. Common determinants of the health status found in the literature are: the family background and genetics (Siest et al., 1993), environmental surroundings, air pollution and community environment friendly attitudes (Capon, 2008), access to infrastructure (Schultz, 2009) and sport (Thorlindsson et al. 1990).

In this paper we approximate human capital endowment by a mixture of education attainment and health. Having rescaled education level and health status variables, we construct a synthetic human capital index ranging from 0 to 1. One stands for a hundred percent human capital endowment and zero represents no human capital possessed at all. The synthetic human capital indicator is modeled as a function of three hidden, unobservable elements which for simplicity we call: health, inborn intelligence, and knowledge acquired through education and training. Treating human capital as latent and developing indicators for selected human capital dimensions may allow to perform quantitative analyses of the phenomenon, such as latent class cluster analysis and latent class factor analysis and to carry out regional and international comparisons.

The assumption that unobservable variables exist and are mathematically related to manifest variables allows an empirical researcher to go beyond his or her dataset and to speak as a theoretician, not merely a statistician (Lazarsfeld and Henry 1968). After Gigerenzer et al. (1989), Henry (1999) postulates that Spearman's concept of intelligence roots from a statistical construct defined by the method of factor analysis - "The Vectors of the Mind".

However, even if we manage to distinguish three latent human capital elements, it becomes clear that these three dimensions are interrelated and have common determinants. An example is access to infrastructure which influences both health state and the capability of knowledge accumulation. Family background affects both inborn intelligence and education attainment. Given the overlaps between the three human capital segments, we decided to apply a simple cross-section model where indicators for each human capital latent segment will be modeled together. This approach could be extended to include three latent variables at later stage of the research.

III. Data description

We use the COMPETE database, which is an independent socio-economic survey covering a wide spectrum of various aspects of life ranging from demographic structure of the household, the economic and the labour market activity, wealth, living conditions and income to the qualitative measures such as involvement in sport and cultural activity or healthcare. The sample is representative, however covers only five selected local communities (gminas) in Poland, namely Gostyń, Gliwice, Manowo, Małogoszcz and Zgierz.

Gostyń represents economically developed local communities (gminas) from the Wielkopolskie. Gliwice lies in the Upper Silesia conurbation, large industrial and post industrial area where mines and steel mills were dominant industries. Manowo represents poor, rural area of the Zachodniopomorskie. Małogoszcz lies in poor rural area of central Poland in the Swietokrzyskie region. Local community (gmina) Zgierz

is a rural area in the suburbs of city of Zgierz and city of Lodz. The diversified sample makes an opportunity to compare the level of social capital in the different regions of Poland.

IV. Model

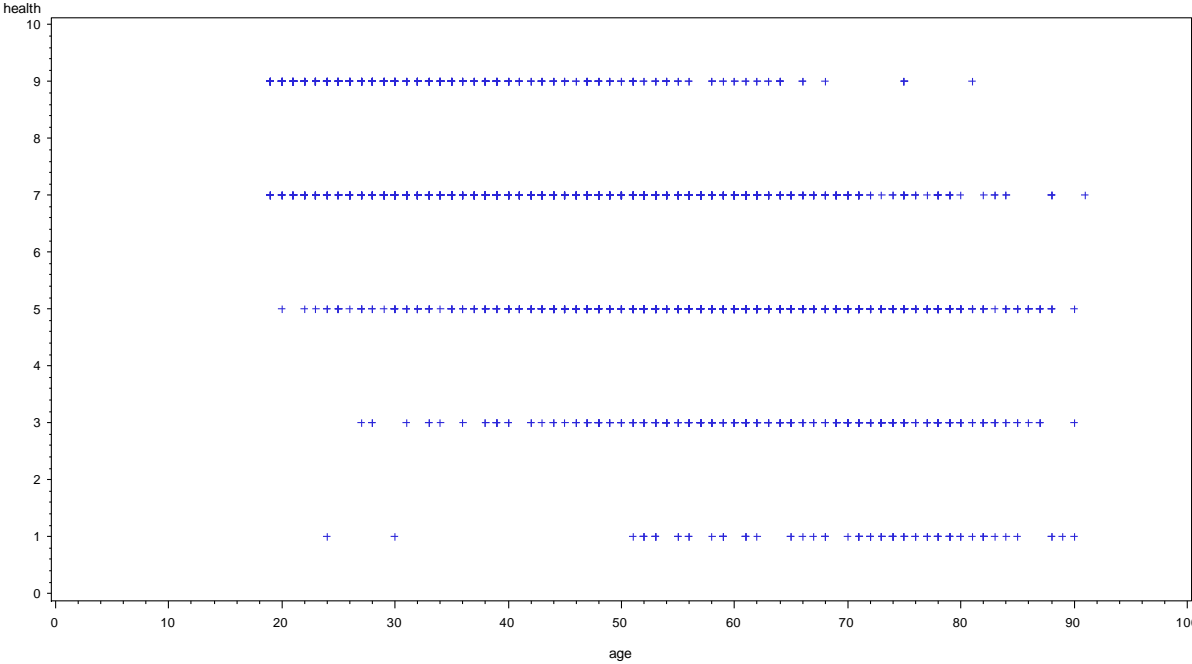
The first step of our estimation procedure is finding manifest variables for each of the three defined elements of human capital: health, inborn intelligence and knowledge acquired through education and training.

Five indicators have been selected for the health status: fact of consuming vegetables regularly (**veg**), doing sports (**sport**), visiting doctors regularly (**doct**), environmental attitudes (**eco**) and smoking (**smoke**).

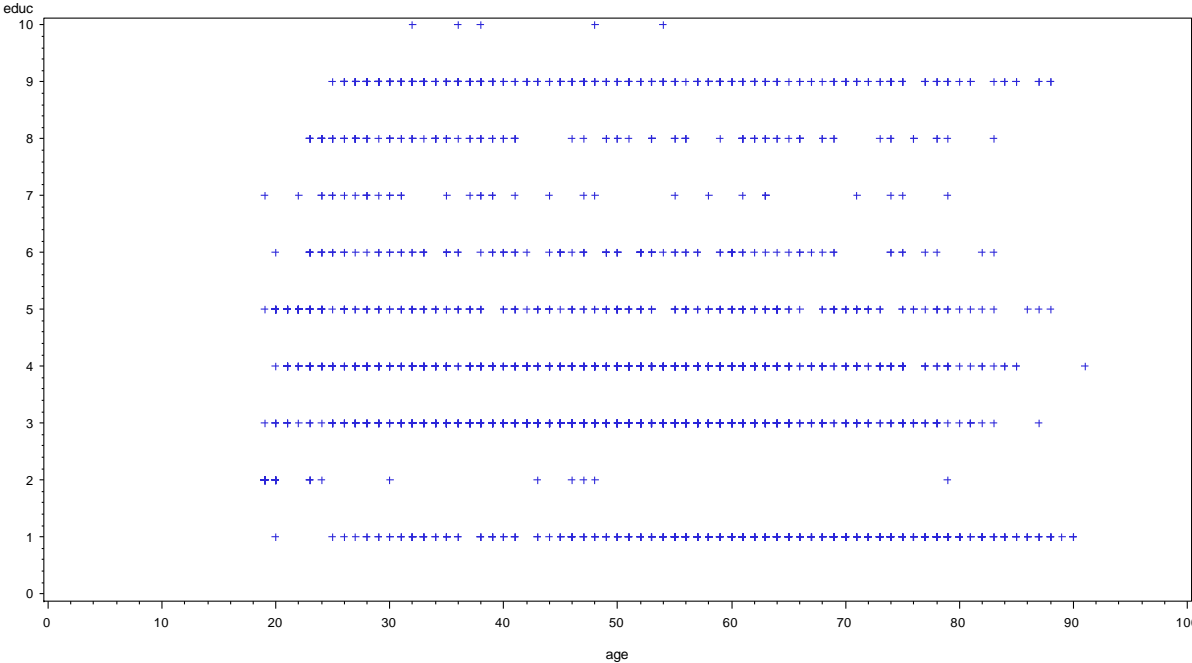
Acquired education has been proxied by respondent's aspirations (**aspchild** - importance of choosing a good-quality school for respondent's children), respondent's ambitions (**amb** – belief in the importance of education in finding a better job in the future), access to physical and virtual infrastructure (**internet** – fact of accessing the internet, **mobile** – fact of possessing a mobile phone, **rural** - fact of living in urban or rural area, **dist** - distance to work), **info** – time spent on watching news on TV, frequency of social contacts (**friends**), willingness to follow trainings at work (**train**), tolerance (**toler**) and civil engagement (**elec** – participation in local elections).

We do not have good proxies for inborn intelligence in our dataset since the information on respondents' family background is very limited. The following indicators have been selected: support from other family members (**famsupp**), **childres** (children's results at school), careful reading of contracts before signing (**contr**), propensity to being self-employed (**selfempl**) and trust to others (**trust**). We extended our model by adding two socio-economic variables: age and gender (**sex**).

Below we present plots of health and education versus age.



Graph 1. Scatter plot of health status versus age



Graph 2. Scatter plot of education levels versus age

We can clearly observe a negative relationship between age and health. The relationship between education and age is more ambiguous. The highest level of education is observed for respondents aged between 30 and 55.

V. Estimation

The dependent variable is a product of education and health index. The education index takes ten different values from 0 to 1:

- 0,1 – primary school education
- 0,2 – junior high school education
- 0,3 – trade school education
- 0,4 – secondary trade school education
- 0,5 – secondary school education
- 0,6 – post-secondary school education
- 0,7 – not completed higher education
- 0,8 – engineering / bachelor studies
- 0,9 – master studies
- 1,0 – PhD studies

The health index takes five different values from 0 to 1:

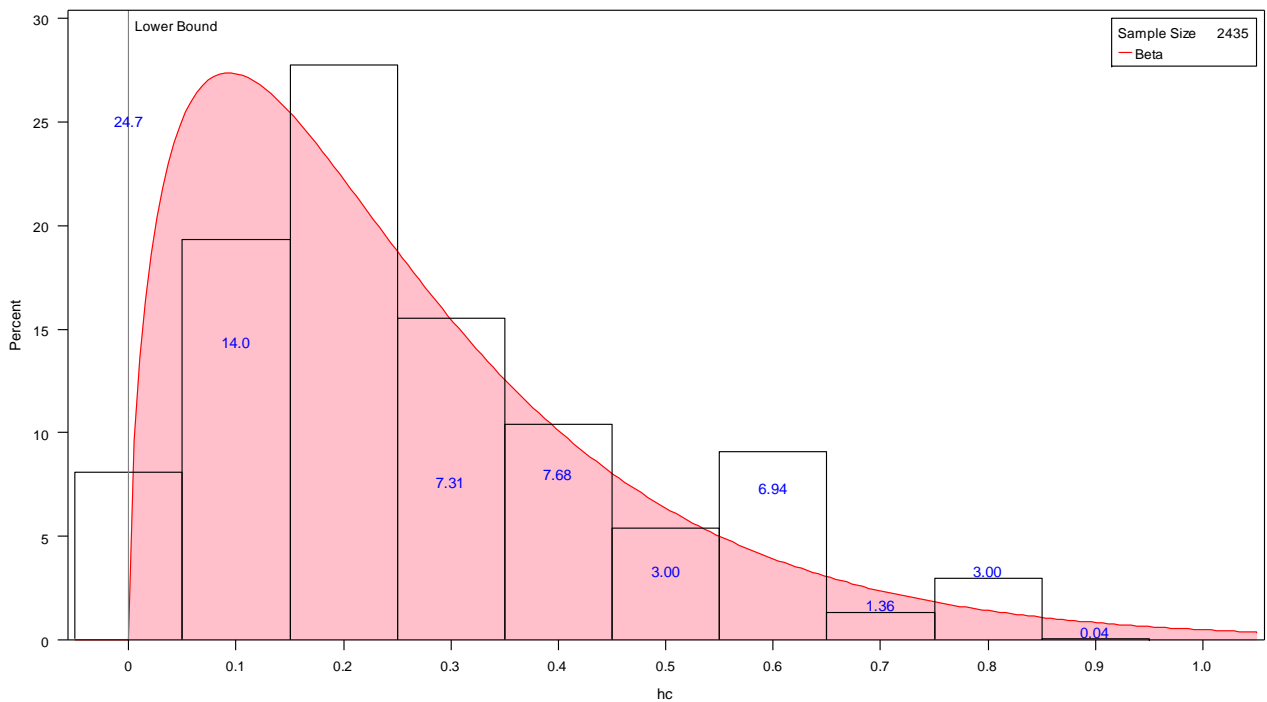
- 0,9 – very good
- 0,7 – good
- 0,5 – not good and not bad
- 0,3 – bad
- 0,1 – very bad

Given that the dependent variable is a fraction, we could so we could not apply a standard linear regression. Due to the floor, non linear effects and heteroscedasticity may arise and errors become not normal. Alternative techniques are: a general squares model or a beta model. Looking at frequency table and the distribution of the dependent variable (graph 3) and fitting it on a graph together with a beta distribution ($\theta=1$ and $\text{scale}=5$) we decided to model the synthetic index of human capital using a beta distribution. The beta distribution is bounded between 0 and 1 (but does not include either 0 or 1).

The FREQ Procedure

ind	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	601	24.68	601	24.68
0.1	341	14.00	942	38.69
0.2	779	31.99	1721	70.68
0.3	178	7.31	1899	77.99
0.4	187	7.68	2086	85.67
0.5	73	3.00	2159	88.67
0.6	169	6.94	2328	95.61
0.7	33	1.36	2361	96.96
0.8	73	3.00	2434	99.96
0.9	1	0.04	2435	100.00

Fitted Beta Distribution of Human Capital



Graph 3. Distribution of the dependent variable.

We allow respondents' human capital endowment to vary depending on different values of the explanatory variables, We express the average human capital index **hc** as a linear function of the following parameters:

$$\mu_i = f(b_0 + b_1*age + b_2*sex + b_3*info + b_4*dist + b_5*supfam + b_6*age*age + b_7*toler + b_8*trust + b_9*amb + b_{10}*aspchild + b_{11}*contr + b_{12}*elec + b_{13}*selfempl +$$

b14*mobile + b15*internet + b16*doct + b17*veg + b18*smoke + b19*sport + 20*friends + b21*eco + b22*train + b23*childres + b24*rural)

To ensure that μ_i remains between 0 and 1 we use the following logistic transformation:

$$\mu_i = \frac{e^{b_0 + b_1 x_1 + b_2 x_2 + \dots}}{1 + e^{b_0 + b_1 x_1 + b_2 x_2 + \dots}}$$

Given the traditional probability density function of the beta distribution with two shape parameters α and β :

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

as well as the alternative parametrization (form corresponding with the conventions of Generalized Linear Models with one location and one scale parameter where the mean distribution of the dependent variable is modeled as a function of the explanatory variables):

$$f(y|\mu, \Phi) = y^{\mu\Phi-1}(y-1)^{(1-\mu)\Phi-1}$$

we calculate α and β expressed with the earlier defined μ and Φ in the following way:

$$\begin{aligned}\alpha &= \mu * \Phi \\ \beta &= \Phi - \mu * \Phi\end{aligned}$$

Following this operation we define the log-likelihood function as:

$$LL = \ln(\Gamma(\alpha+\beta)) - \ln(\Gamma(\alpha)) - \ln(\Gamma(\beta)) + (\alpha - 1)*\ln(hc) + (\beta - 1)*\ln(1 - hc)$$

and we estimate the model using non-linear maximum likelihood procedure. Below the estimated parameters are presented.

Two-Way Layout
The NL MIXED Procedure
Specifications

23:23 Tuesday, July 27, 2010

Data Set LIBRARY.COMPETE3
Dependent Variable hc
Distribution for Dependent Variable General
Optimization Technique Trust Region
Integration Method None

Dimensions

Observations Used 657
Observations Not Used 1815
Total Observations 2472
Parameters 25

Two-Way Layout
The NL MIXED Procedure

23:23 Tuesday, July 27, 2010

NOTE: GCONV convergence criterion satisfied.

Fit Statistics

-2 Log Likelihood -521.8
AIC (smaller is better) -475.8
AICC (smaller is better) -474.1
BIC (smaller is better) -372.6

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
const	0.3458	2.7003	662	0.13	0.8981	0.05	-4.9563	5.6479	-0.00147
age	-0.00957	0.002889	662	-3.31	0.0010	0.05	-0.01524	-0.00390	0.413635
gender	0.1307	0.06355	662	2.06	0.0401	0.05	0.005924	0.2555	-0.00203
info	0.8797	0.1564	20	5.63	<.0001	0.05	0.5535	1.2058	-0.35933
dist	-0.9140	0.2173	20	4.21	0.0004	0.05	0.4606	-0.70876	1.3673
suppfam	1.0007	0.2865	20	3.49	0.0023	0.05	0.4030	1.5984	-1.07758
toler	0.01361	0.01399	662	0.97	0.3311	0.05	-0.01386	0.04108	-0.00141
trust	-0.04258	0.01780	662	-2.39	0.0170	0.05	-0.07754	-0.00763	-0.00197
amb	-0.1756	0.07898	662	0.55	0.5855	0.05	-0.1120	-0.0012	-0.9812
aspchild	0.04309	0.06177	662	-2.84	0.0046	0.05	-0.2969	0.00052	0.05428
readcontr	-0.03970	0.02565	662	-1.55	0.1222	0.05	-0.09007	0.01067	-0.00172
elect	-0.3388	0.05979	662	-5.67	<.0001	0.05	-0.4562	-0.2214	-0.00168
selfempl	0.07924	0.09048	662	0.88	0.3815	0.05	-0.09843	0.2569	-0.00228
mobile	0.03104	0.3143	662	0.10	0.9214	0.05	-0.5862	0.6483	-0.00152
internet	0.3004	0.09658	662	-3.11	0.0019	0.05	0.1107	0.4900	-0.00151
doct	0.3478	0.06426	662	5.41	<.0001	0.05	0.2216	0.4740	-0.00188
veg	-0.02254	0.03729	662	-0.60	0.5458	0.05	-0.09576	0.05069	-0.00226
smoke	-0.2161	0.06051	662	-3.57	0.0004	0.05	-0.3349	-0.09729	-0.00073
sport	-0.3185	0.06311	662	-5.05	<.0001	0.05	-0.4424	-0.1946	-0.00183
friends	0.4509	0.05940	20	7.59	0.0038	0.05	0.3270	0.5748	-1.78502
eco	-0.07364	0.06954	662	-1.06	0.2900	0.05	-0.2102	0.06291	-0.00172
train	0.03503	0.004836	20	7.24	<.0001	0.05	0.02494	0.04512	-9.78011
childres	1.5490	0.2518	20	6.15	<.0001	0.05	1.0237	2.0743	-0.62856
rural	-0.4056	0.3504	20	-1.16	0.2606	0.05	-1.1365	0.3253	-0.94958

Overall, the model is significant and has desirable statistical properties. The signs of most coefficients are in accord with our expectations and the economic theory. Regular access to health infrastructure and visiting doctors had a positive impact on the respondent's human capital. In line with intuition, smoking has a negative impact on health and consequently on human capital. The statistically insignificant coefficients for vegetables consumption and environmental attitudes mean that the effect of these variables on the health status is too weak to become noticeable.

Contrary to expectations, doing sport was found to have a negative influence on human capital. An explanation may be that its positive impact on the health element is outbalanced by negative impact on education and training. Doing sport takes time so a substitution effect may dominate resulting in an overall negative effect of doing sport on human capital.

Respondent's aspirations, proxied by the importance of choosing a good-quality school for respondent's children, access to information infrastructure, expressed by the fact of using the internet or watching news on TV and respondent's willingness to follow trainings all had positive impact on respondent's human capital. A statistically significant and positive coefficient for social contacts with friends means that social networks improve individual capital endowment either by the health or education channel. The coefficient for possessing a mobile phone was not significant which may be due to the fact that this effect is already encapsulated in the coefficient for internet usage. Distance to work had a negative effect on human capital accumulation, however, the fact of living in a rural area did not have a significant impact.

As far as indicators for the inborn intelligence are concerned, both family support and children's results at school had a positive effect on human capital. A negative coefficient for age indicates that human capital decreases with age. Although normally work experience has increasing returns on human capital, this effect is outbalanced by the negative effect of age on health, the latter being the dominant effect. The positive and statistically significant coefficient for gender indicates that

male respondents possess on average higher human capital than female respondents.

VI. Conclusion

In our research we took the advantage of using a unique dataset containing a wide range of social and demographic measures at the local community (gmina) level to explore the determinants of human capital, expressed as a combination of health and education. On the basis of the available literature, we defined three unobservable, latent human capital dimensions: health, inborn intelligence and acquired skills. From the wide spectrum of sociological, psychological and economic variables we selected indicators for each of the three human capital segments and we constructed a cross-section beta model where we analyzed the impact of individual factors on human capital accumulation.

Overall, the results of our model are in line with expectations and human capital theory. An interesting observation was the decreasing return of age on human capital. This finding can be explained by the fact that work experience, normally increasing with age, is outbalanced by negative effect of age on health, the latter being the dominant effect.

As can be expected, social networks contributed positively to individual human capital endowment. Surprisingly, other social factors such as participation in local elections and trust were found to have a negative impact on human capital. An explanation for this effect may be that better educated persons are more aware of the dangers surrounding them and therefore are more prone to remain skeptical and to distrust others.

Bibliography

Becker, G. (1975) Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, National Bureau of Economic Research.

Buis, M. (2005) Proportions as Dependent Variables, Vrije Universiteit Amsterdam, Department of Social Research Methodology.

Capon, A. (2008) Investigating Health Effects in a Community Surrounding, Environmental Health 2008.

Fleischhauer, K. (2007) A Review of Human Capital Theory: Microeconomics, Discussion Paper no. 2007-01.

Grochowska, A. (2009) Use of Latent Class Models in Modelling Civilization Competences, in: Civilizational competences and regional development in Poland, Z. B. Liberda, A., Warsaw University Press, Warsaw, pp. 33-46.

Grossman M. (1972) On the Concept of Health Capital and the Demand for Health. The Journal of Political Economy. Vol 80, No. 2, pp. 223-255.

Grossman M. (2001) Household Production and Health, City University of New York Graduate Center and National Bureau of Economic Research, Review of Economics of the Household, pp. 331-342.

Henry, N. (1999) Latent Structure Analysis at Fifty. Virginia Commonwealth University, Richmond, Working Paper.

Lazarsfeld, P., and N. Henry (1968) Latent Structure Analysis. Boston: Houghton.

Liberda B., M. Peczkowski (2009) Regional Distribution of Household Income and Household Human Capital in: Civilizational competences and regional development in Poland, Z. B. Liberda, A. Grochowska, eds., Warsaw University Press, Warsaw, pp. 105-130.

Melody, W. (1999) Human Capital in Information Economies, New Media & Society, no. 1, 39-46.

Mincer, J. (1958) Investment in Human Capital and Personal Income Distribution, The Journal of Political Economy, University of Chicago Press, vol. 66.

Mushkin S. (1962) Health as an Investment. The Journal of Political Economy, Vol 70, No. 5, Part 2: Investment in Human Beings, pp. 129-157.

Schultz, P. (2010) Population and Health Policies, Department of Economics, Yale University, Box 208269, New Haven, Connecticut 06520-8269, USA.

Siest, G. et al. (1993) The Role of Genetics in defining Reference Values and Health Status, *Methods Inf Med*, 1993 Apr, 32(3), 255-9.

Thorlindsson T., Vilhjalmsón R., Valgeirsson G. (1990) Sport Participation and Perceived Health Status: a study of Adolescents, Department of Social Sciences, University of Iceland, Odda, Reykjavik, *Social Science and Medicine*, 1990;31(5):551-6.