STATISTIKA

STATISTICS
AND ECONOMY
JOURNAL

VOL. **49** (4) 2012

CZECH
STATISTICAL
OFFICE

EDITOR-IN-CHIEF

EDITORIAL BOARD

ASSOCIATE EDITORS

MANAGING EDITOR

CONTENTS

ANALYSES

INFORMATION

About Statistika

The journal of Statistika has been published by the Czech Statistical Office since 1964. Its aim is to create a platform enabling national statistical and research institutions to present the progress and results of complex analyses in the economic, environmental, and social spheres. Its mission is to promote the official statistics as a tool supporting the decision making at the level of international organizations, central and local authorities, as well as businesses. We contribute to the world debate and efforts in strengthening the bridge between theory and practice of the official statistics. Statistika is a professional journal included in the list of peer-reviewed scientific periodicals published in the Czech Republic and rated by the methodology of the Council for Research, Development and Innovation of the Government of the Czech Republic. Since 2011 Statistika has been published quarterly in English only.

Publisher

The Czech Statistical Office is an official national statistical institution of the Czech Republic. The Office main goal, as the coordinator of the State Statistical Service, consists in the acquisition of data and the subsequent production of statistical information on social, economic, demographic, and environmental development of the state. Based on the data acquired, the Czech Statistical Office produces a reliable and consistent image of the current society and its developments satisfying various needs of potential users.

Contact us

Journal of Statistika | Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czech Republic
e-mail: statistika.journal@czso.cz | web: www.czso.cz/statistika_journal

Application of Regional Price Levels on Estimation of Regional Macro-Aggregates Per Capita in PPS

Petr Musil[1] | *University of Economics, Prague, Czech Republic*
Jana Kramulová[2] | *University of Economics, Prague, Czech Republic*
Jan Čadil[3] | *University of Economics, Prague, Unicorn College, Prague, Czech Republic*
Petr Mazouch[4] | *University of Economics, Prague, Czech Republic*

Abstract

Comparison of Czech regions is almost always based on nominal indicators, i.e. an assumption of the same price levels is included in the analysis, though it is obvious that this assumption is quite strong or even not fulfilled. Results of such an analysis could be seriously affected and simultaneously lead to misinterpretation. The aim of this paper is to estimate regional price levels in the Czech Republic NUTS 3 regions and to estimate their impact on regional macro-aggregates such as the net disposable income. Eurostat uses for this issue the methodology introduced in the common OECD and Eurostat Manual on purchasing power parities (PPP) resulting in purchasing power standard (PPS) estimation at international level. From the theoretical point of view PPS could be estimated also at the regional level, i.e. each region would be considered as an "independent unit". In practice, in most EU countries the problem with poor regional data availability arises. The official methodology is used in this paper with several modifications described. Finally macro indicators adjusted to regional price levels are compared with the "original" data.

Keywords

Czech NUTS 3 regions, price levels, GDP, net disposable income, EKS method, PPS

JEL code

E31, R10

INTRODUCTION

One of the important economic and statistical issues is to study the economic conditions in the different parts of the world. When comparing countries worldwide according to economic performance, usu-

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic. Corresponding author: e-mail: petr.musil@vse.cz, phone: (+420)224095435. Author is working also at the Czech Statistical Office, Na padesátém 81, 100 82 Prague, Czech Republic, e-mail: petr.musil@czso.cz.
[2] University of Economics, Prague, Faculty of Economics, Nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic.
[3] University of Economics, Prague, Faculty of Economics, Nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic; Unicorn College, V Kapslovně 2767/2, Prague 3, Czech Republic.
[4] University of Economics, Prague, Faculty of Informatics and Statistics, Nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic.

ally GDP per capita in purchasing power standard (PPS) is used. But estimating of PPS in the European Union is based on average prices in countries that means on national prices, therefore local differences within one country are not taken into account. This causes problems and expected inaccuracy, when we want to compare regions. Recently the attention has moved from observation of international differences to interregional differences. Also the question of regional convergence or divergence stands at the forefront of the economic analyses.

That is why we are focusing on the topic of regional price levels. Surprisingly this topic has not drawn any attention in the Czech Republic so far. We will below briefly discuss the situation in some other countries.

The main aim of this paper is to estimate regional price levels for the Czech Republic NUTS 3 regions, the partial aims are then to estimate some of the main macroeconomic indicators for Czech NUTS 3 regions and then to appraise the results obtained. The main contribution of this paper is the adjustment of the official Eurostat / OECD methodology that is more suitable for national level. We decided to work with the level of 14 NUTS 3 regions, because that is according to our opinion the lowest level for which the data needed are available in sufficient extent. But even at this level we were facing several problems when searching for some data.

This paper is divided into several sections. Firstly we perform a short overview of development of economic comparison at the national level and introduce several attempts of regional price levels estimates in different counties. In section 2 the official Eurostat / OECD methodology is described as well as our adjustments needed to be done for the comparison at the regional level. After that the most important data sources are introduced and finally the obtained results are discussed.

1 IMPORTANCE OF THE TOPIC

Comparison of price levels has been an interesting and important topic since 1960s, when the International Comparison Program (hereinafter ICP) was established in 1968 (see International Bank for Reconstruction and Development / The World Bank, 2008). While in 1970 only 10 countries were involved in this program, in 2011 it was already around 200 countries. Table 1 shows both regional dispersion and income dispersion of the countries involved in this program in 2011.

Table 1 International Comparison Program 2011 – involved countries

ICP region group / Income group	High	Upper middle	Lower middle	Low	Total
Africa	1	8	15	28	52
Asia & Pacific	5	2	11	5	23
Commonwealth of Independent States	0	4	4	2	10
Latin America & Caribbean	9	21	10	1	41
OECD – Eurostat	35	12	0	0	47
Pacific Islands	1	3	9	2	15
Western Asia	6	1	7	0	14
Singleton Countries	0	1	2	0	3
Dual Participation	x	x	x	x	–5
Total	**x**	**x**	**x**	**x**	**200**

Source: International Comparison Program (2011), authors' adaption

Apart from ICP also ECP (European Comparison Program) exists, which is the regional program for Europe. Czech Republic (CZ) was involved in 1993 (comparison CZ and EU countries based on bilateral comparison between CZ and Austria) and in 1996 (multilateral comparison). Since 1999 Czech Republic has been a regular part of this yearly comparison (Czech Statistical Office, 2012). Table 2 shows 47 countries included into this program divided into five groups (four of them constituted by Eurostat and one by OECD). Each Eurostat group has its group leader.

Table 2 European Comparison Program 2011 – five groups of countries with group leaders

Eurostat				OECD
Northern group	Western group	Eastern group	Southern group	OECD
Finland	**Netherlands**	**Austria**	**Portugal**	Australia
Denmark	Belgium	Bosnia and Herzegovina	Albania	Canada
Estonia	Czech Republic	Bulgaria	Cyprus	Chile
Iceland	France	Croatia	FYR of Macedonia	Israel
Latvia	Germany	Hungary	Greece	Japan
Lithuania	Ireland	Montenegro	Italy	Korea
Norway	Luxembourg	Romania	Malta	Mexico
Poland	Switzerland	Serbia	Spain	New Zealand
Sweden	United Kingdom	Slovakia	Turkey	Russian Federation
		Slovenia		United States

Note: Group leaders in bold.
Source: Eurostat-OECD Methodological Manual on Purchasing Power Parities (PPPs), 2012 (preliminary version), authors' adaption

In this ECP program countries are multilaterally compared on the basis of Éltetö-Köves-Szulc (EKS) method (brief explanation of the method procedure see below, for more information see Jílek, Moravová, 2007, pp. 227–229, or European Commission, 2006).

Apart from the national comparisons, also attempts of regional comparisons of price levels exist. This effort can be seen e.g. in Great Britain (Hayes, 2005), United States of America (Aten, D'Souza, 2008), China (Brandt, Holz, 2006) or very often Germany (Blien et al., 2009, Dreger et al., 2010, Roos, 2006). In the Czech Republic, as it was already mentioned, no attention has been paid to this topic so far. Lower level of discussion of this topic can be in the fact, that for analysis of this type, a big amount of data is necessary. That may be especially in larger countries quite difficult.

Importance of this topic can be seen not only in the endeavour to measure economic phenomenon properly, but also in consequences that arise from decisions based on regional indicators' values. These are e.g. regional policy decisions that are usually based on GDP per capita in PPS. Also the convergence / divergence discussions (see Čadil, Mazouch, 2011) usually come out of regional or national GDP per capita in PPS.

The PPP / PPS (Purchasing Power Parity / Purchasing Power Standard) is computed in accordance with the methodology formed by OECD and EUROSTAT. This methodology is introduced in Section 2.

2 METHODOLOGY

2.1 Eurostat / OECD Methodology

Comparison of economic development of countries was formerly based on the conversion of macroeconomic indicators to widespread currency (usually US dollar) by the exchange rate. This method is relatively simple, but the exchange rate is affected by the currency's supply and demand, intervention of central bank, speculation etc. Moreover, some services cannot be traded (e.g. defense, public administration) so the coverage of this approach is limited to negotiable products. Therefore OECD and Eurostat developed the PPP methodology that eliminates these problems. Artificial currencies (units) were introduced which have the same purchasing power in all involved countries: PPS (purchasing power standard) for comparison of EU countries and OECD dollar for OECD countries. Macroeconomic indicators expressed in PPS or OECD dollar can be easily compared and resulting differences comfortably assessed.

This comparison can be done at any level of aggregation. Expenditure approach is used in the calculation. Each component of GDP is divided into so called "basic headings" that represent minimum level for which expenditure weights are available. Each member state is supposed to choose products that are representative for every single basic heading (at least one product per basic heading). In addition to this member countries have to collect prices of representative products as well as prices of products that are

not representative but available at domestic market, however they are representative in other country. Otherwise international comparison could not be done. Data collection can be done within the whole country or in the capital city only. In the second case, which is more frequent, countries are supposed to provide spatial adjustment factors to obtain national prices.

The procedure of PPPs calculation can be described in 6 steps stated below. This procedure must be done for any product heading. The methodology is in detail described by European Commission (EC, OECD, 2006, Annex V).

The first step is the calculation of a Laspeyres type PPPs matrix. In general, Laspeyres type of price and volume indices uses weights from basic period. In the term of PPPs calculation, where we compare space instead of time, Laspeyres index refers to the base country. The computation formula (1) of Laspeyres index (country *B* to country *A*) for each basic heading can therefore be expressed as follows:

$$L_{B/A} = \left[\prod_{i=1}^{n} \frac{p_{B_i}}{p_{A_i}} \right]^{\frac{1}{i}}, \tag{1}$$

where *p* is a price of a certain product. It is important to add, that only products that are representative for country *A* are taken into account. This calculation is done for all countries in order to obtain Laspeyres type PPPs matrix.

The second step represents computation of a Paasche type PPPs matrix. Generally Paasche type of index operates with weights from current period, in PPPs calculation Paasche index refers to the partner country. Paasche index (country *B* to country *A*) is calculated according to the following formula (2):

$$P_{B/A} = \left[\prod_{i=1}^{n} \frac{p_{A_i}}{p_{B_i}} \right]^{\frac{1}{i}} = \frac{1}{L_{A/B}}, \tag{2}$$

in this case only prices of products that are representative for partner country are included in the calculation. Due to a relation between Laspeyres and Paasche indices Paasche PPPs matrix can be simply completed by using Laspeyres indices what facilitates the calculation.

The third step consists in the calculation of a Fisher type PPPs matrix as a geometric mean of corresponding Laspeyres and Paasche indices. In general, Fisher indices are reversal (3) but not transitive (4).

$$F_{B/A} \times F_{A/B} = 1, \tag{3}$$

$$F_{B/A} / F_{C/A} \neq F_{A/C}. \tag{4}$$

The fourth step stands in completing of the Fisher type PPPs matrix. Actually the problem can occur as the computed matrix can be incomplete due to missing prices. Product that is representative in the base country may not be available in the partner country and therefore the price in this country does not exist. For estimating the missing indices the procedure of so called bridging can be used, that means another country is used as a bridge. For example in equation (5) $F_{A/B}$ cannot be calculated directly, but it can be estimated when countries *C* and *D* are used as a bridge:

$$F_{A/B} = \left(\frac{F_{A/C}}{F_{B/C}} \frac{F_{A/D}}{F_{B/D}} \right)^{\frac{1}{2}}. \tag{5}$$

Generally missing index is estimated as a geometric mean of all the indirect indices.

The fifth step lies in the calculation of EKS PPPs matrix. EKS (already mentioned Éltetö-Köves-Szulc) method is used in order to estimate transitive indices. EKS PPPs are calculated as an unweighted according to formula (6):

$$EKS_{B/A} = \left[\prod_{i=A}^{n} \frac{F_{B/i}}{F_{A/i}} \right]^{\frac{1}{\sum n}}. \quad (6)$$

The final, sixth, step, is standardization of EKS PPPs matrix. Equation (7) shows the standardisation, provided by a joint basis when a price of one basic heading of one region is related to all other countries:

$$EKS_A = \frac{EKS_{A/A}}{\left[\prod_{i=A}^{n} EKS_{i/A} \right]^{\frac{1}{\sum i}}} = \frac{EKS_{A/B}}{\left[\prod_{i=A}^{n} EKS_{i/B} \right]^{\frac{1}{\sum i}}} = \ldots = \frac{EKS_{A/n}}{\left[\prod_{i=A}^{n} EKS_{i/n} \right]^{\frac{1}{\sum i}}}. \quad (7)$$

This calculation of PPPs is done for each product heading that represents the minimum level of aggregation for which calculation is done. PPPs can be aggregated by Laspeyeres type of PPPs that uses weights from base country or Paasche type of PPPs which is based on weights from partner country. Fisher type of PPPs is their geometric mean. EKS method is used in order to ensure transitivity, than standardisation procedure is employed.

2.2 Adjustments of the official methodology for Czech regions comparison

Our estimate of regional price levels in Czech NUTS 3 regions is inspired by just described PPP methodology. Comparison of regions within one country is expected to be similar to comparison of countries, but several differences were indentified. At first less data on regions than on countries is available. At national level data on production, expenditure and income approach is available however at regional level only limited data usually exists. EU countries are not obliged to transmit figures about expenditure approach and therefore the Czech Republic does not compile regional GDP from expenditure side at all (just output and income approaches are regularly published). Generally the main difficulty represents the foreign trade field because export and import from/to all regions would have to be estimated. Custom systems Intrastat and Extrastat, which are the main data sources for external trade at the national level, are not useful at regional level. Therefore we decided to base our estimate on final household consumption which is the main component of expenditure approach (covers approx. 50%) and for which substantial differences are expected. Other components of GDP, such as gross (fixed) capital formation or government consumption are not included into the computation.

As was already stated, PPP methodology is used with several adjustments. First adjustment consists in the level of calculation. According to PPP methodology the minimum level is so called basic heading. Instead of the level of basic product headings (about 148 items for household consumption) we performed our computation at the level of representatives (about 700 items) which represents significantly higher level of disaggregation and is much more detailed than product headings breakdown. Another adjustment refers to handling with missing data. These are estimated in a different way. Although PPP methodology employs bridging i.e. other countries are used as a bridge, different method is used in our case. When missing prices occur, this usually means that product is not available in the particular region and citizens have to purchase the specified product in any other region (e.g. eye surgery). Missing prices are in our approach estimated as an arithmetic average of prices in other regions. All products are supposed to be representative in all regions as regions within country are more similar to each other than countries (the same currency, similar shopping manners etc.). Consequently Laspeyres, Paasche and Fisher type of PPPs are the same and all indices are transitive.

Concerning the structure of household expenditure we can mention one more adjustment. As already stated, estimate of regional PPP is based on the level of representatives (about 700). Final household consumption expenditures are available in the classification CZ COICOP (Classification of Individual Consumption According to Purpose) at 4-digit level as the most detailed level of aggregation. More

detailed data can be found in household budget survey only. Calculation is based on national accounts data; within CZ COICOP 4-digit level linear interpolation is used in order to obtain weights at the level of representatives. It is clear that the structure of household consumption is not completely the same in all regions. Therefore regional structure has to be estimated. In comparison with official methodology different concept of household consumption is employed. PPP methodology is based on domestic concept which includes all expenditures of households in domestic economy regardless a purchase is made by resident or non-resident whereas "our" regional consumption basket is based on national concept. The main reason of this is on one hand a scarce of data on domestic concept, on the other hand certain regionalization of data in national concept has been done in academic environment. As rents and selected services for which main differences in prices are expected are not negotiable among regions, final results may not be affected significantly.

3 DATA SOURCES

Fulfilment of the data matrix proved to be more difficult than we expected at the beginning of our research. We needed to search for the data in different sources, including research estimates. All the data are stage by stage being made more and more precise.

National Statistical Institutes are worldwide supposed to provide data for PPP programme. Data on weights are derived from annual national accounts. Concerning prices countries have a choice to collect prices of all products within a single year or to collect data over three years. The first option is quite resource-demanding therefore most countries prefer the second possibility which means that consumer basket is divided into six parts. Every half a year one sixth of prices is surveyed. Prices of remaining two thirds of products that are not surveyed that year are estimated by using temporal adjustment factors. There are two opportunities how to ensure that all prices will be representative for the whole country. The first one is that the data are collected in the capital city and adjusted by given spatial factors. Second possibility expects that the data collection is not limited to capital city. It is obvious that the surveyed stores should be selected with respect to shopping manners in each country (or region). In comparison to the representatives' sample for consumer price index the description of them in this case is much more detailed, because it is crucial for the comparability of countries (or regions).

Calculation of regional price level is based on several data sources. The main data source in our case is the consumer price survey that is in the Czech Republic conducted monthly. Generally this survey replaces PPP survey because of following reasons. More detailed data on weights and prices are available in consumer price survey (about 700 representatives in comparison with 148 basic heading in PPP survey). There is no need to use temporal adjustment factors because survey is carried out monthly. Annual average prices are used in order to eliminate possible swings in monthly data. But unfortunately not the whole consumption basket is covered by this data source so we were forced to use also other ones. For example data on paid rents were provided by Institute for Regional Information (hereinafter IRI) that collects data on rents at very detailed regional stratification (263 territorial units within the Czech Republic). In this survey a so called model flat is defined (that has same flat dimensions, age, level of depreciation etc.) and the prices for this model flat are collected in all regions. It means that differences in the regional structure of housing fund are not taken into account. Imputed rent is included in the model as well. For each region average imputed rent per m2 is calculated and it is considered as price of living in own occupied dwelling. Web data sources and experts' estimates were also used. Because of data availability we chose 2007 as the year of this analysis. Due to very complicated and long-lasting searching for data and especially long-lasting data matrix completing it was so far not possible to establish a longer time series.

4 RESULTS

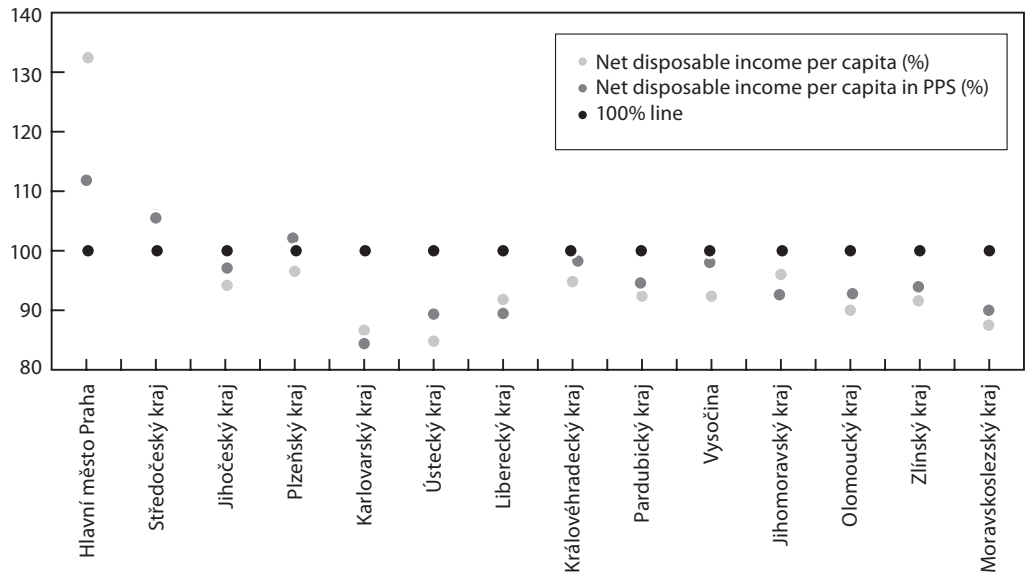
Final household consumption expenditures represent the main use of net disposable income (hereinafter NDI) of households (about 95%). Net disposable income per capita is sometimes considered as an indicator of well-being. Table 3 shows preliminary results for 2007.

Table 3 Comparison of results obtained by official and adjusted methodology (2007)

Region	PPP	Net disposable income per capita	Net disposable income per capita in PPS	Net disposable income per capita (%)	Net disposable income per capita in PPS (%)
Hlavní město Praha	119.7	230 578	192 703	132.2	110.5
Středočeský kraj	101.9	187 150	183 697	107.3	105.4
Jihočeský kraj	97.9	168 100	171 690	96.4	98.5
Plzeňský kraj	97.1	172 868	178 062	99.1	102.1
Karlovarský kraj	101.4	156 050	153 909	89.5	88.3
Ústecký kraj	94.9	152 960	161 171	87.7	92.4
Liberecký kraj	101.4	162 996	160 690	93.5	92.2
Královéhradecký kraj	96.4	168 919	175 158	96.9	100.5
Pardubický kraj	98.2	165 325	168 414	94.8	96.6
Vysočina	95.6	165 652	173 252	95.0	99.4
Jihomoravský kraj	103.4	171 168	165 462	98.2	94.9
Olomoucký kraj	96.9	160 623	165 776	92.1	95.1
Zlínský kraj	100.8	168 523	167 143	96.7	95.9
Moravskoslezský kraj	96.7	157 100	162 533	90.1	93.2
Czech Republic	**100.0**	**174 360**	**174 360**	**100.0**	**100.0**

Source: Authors' calculation

Figure 1 Net disposable income per capita in PPS (in %) with one national price level and regional price levels (2007)



Source: Authors' calculation

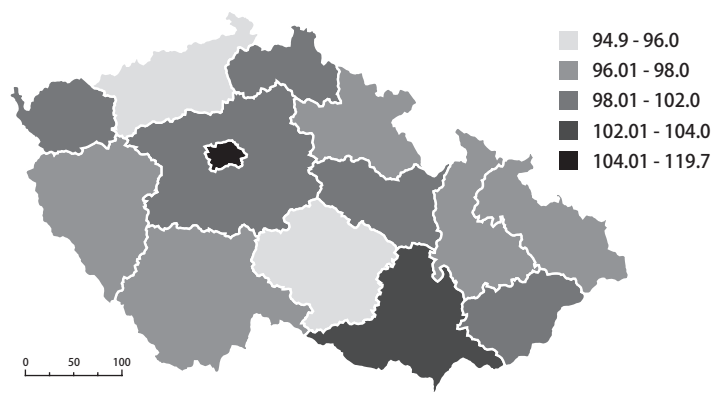
The highest price level is located in the capital city as it was expected which is caused mainly by higher prices of living and some other services (e.g. transportation, personal services). On the other hand it can be marked as a surprise that relatively high price level is in small regions that are not considered as highly developed (Liberecký kraj or Zlínský kraj). Reason for this can be uneven size of the regions according to classification CZ-NUTS 3. In these small regions regional centre (town), where prices are usually higher than in the countryside, play more important role than in large regions.

Our main finding is that differences between regional indicators adjusted to regional price levels are smaller than between the "original" indicators; nevertheless Praha is still the richest region. On the other hand the advantage of Praha decreases just to 10.5%. Figure 1 shows these results. We can see, that the main decrease in NDI in percentage points is in the case of Praha, on the other hand the main increase is in Ústecký kraj.

Originally calculation and usage of PPS is based on GDP however at regional level inappropriate interpretation can lead to misunderstanding. Regional GDP represents regional gross value added (GVA) produced in region plus net taxes on products regardless inputs used in production. Compensation of employees who work in Prague but live in other regions belongs to Praha's GDP though it can be spent in other regions as well. This is probably the main reason for difference between the share of Prague on GDP per capita and share of Prague on net disposable income per capita (compare Chlad et al., 2009). Regional economies can be compared from the point of view of regional GDP, but comparison based on GDP per capita is at least questionable. Moreover place of living in population census in some countries (not in the Czech Republic) is based on administrative data sources (see Šanda, 2012). Some people have corresponding address at place where they actually live, but their official address is elsewhere (moreover place of birth). Sometimes GDP per capita is considered as an indicator of well-being though in this consideration there are many reservations (see Stiglitz et al., 2009).

On Figure 2 the regional dispersion of regional price levels is depicted. As it was stated, highest price levels can be found in regions with two biggest cities, the lowest are in Zlínský kraj, kraj Vysočina and also in Ústecký kraj, which is in a long-term sight affected by structural changes and high unemployment.

Figure 2 PPPs in the Czech NUTS 3 regions in % (2007)



Source: Authors' calculation

CONCLUSION

In this paper we tried to introduce the topic of regional price levels. PPS methodology used for international comparisons is based on one national price level. However, this is no more useful for regional comparisons. We showed that when OECD / Eurostat methodology is adjusted, the results obtained may be different, especially for such regions as Praha or its agglomeration region. Regional indicators can be recalculated in accordance to adjusted methodology and after that compared with the original ones. We showed in this paper as an example the net disposable income that was adjusted to local price level. This approach as we consider should provide more reliable data on living conditions in regions. The results while taking regional price levels into account converge a lot (e.g. NDI in PPS).

We performed our computations for the year 2007 which was suitable according to available data sources. This issue (of data sources) also makes it difficult to obtain a longer time series, which would be important for intertemporal comparison and especially for assessment of regional price levels development, but it is so far almost impossible to construct this time series. Another problem connected with this is, that year-on-year differences could be expected in the extent of statistical error.

This issue is important not only from the point of view of statistics and its accuracy but the results can also potentially be used to adjust regional policy decisions. In fact, these are very often based on GDP per capita in PPS that may be biased when taking into account just one national price level for all the regions.

Possibilities for broadening of our analyses and questions for further research we see especially in the computation of some other years to perform later possibly at least basic comparison in time. On the other hand changes in methodologies and survey techniques make this quite challenging. For example data for the year 2001 and before don't exist at all.

Regional price indices (and especially consumer price index CPI) represent other important related issue that would be desired by local governments for regional policy as well as public. Development of regional price levels in time is determined by changes in price and structure of regional macro-aggregates.

Interesting would definitely be also international comparison, i.e. introduction of this adjusted methodology to other countries to see the differences in results there.

ACKNOWLEDGEMENTS

This paper was supported by the University of Economics, Prague – Internal Grant Agency; Project No. IGA 12 / 2011 "Estimation of regional price levels".

References

ATEN, B. H., D'SOUZA, R. J. Research Spotlight: Regional Price Parities Comparing Price Level Differences across Geographic Areas. *Survey of Current Business*, 2008, 11 (November), pp. 64–74.

BLIEN, U., GARTNER, H., STUBER, H., WOLF, K. Regional Price Levels and the Agglomeration Wage Differential in Western Germany. *Annals of Regional Science*, 2009, 1 (March), pp. 71–88.

BRANDT, L., HOLZ, C. A. Spatial Price Differences in China: Estimates and Implications. *Economic Development and Cultural Change*, 2006, 1 (October), pp. 43–86.

ČADIL, J., MAZOUCH, P. PPS and EU Regional Price Level Problem. *The Open Political Science Journal*, 2011, 4 (November), pp. 1–5.

DREGER, CH., KOSFELD, R. Do Regional Price Levels Converge? *Jahrbücher für Nationalökonomie und Statistik*, 2010, 3 (June), pp. 274–286.

EC, OECD. *Eurostat-OECD Methodological Manual on Purchasing Power Parities*. Paris: European Communities / OECD, 2006.

EC, OECD. *Eurostat-OECD Methodological Manual on Purchasing Power Parities (PPPs)*. Paris: European Communities / OECD, 2012 (preliminary version).

HAYES, P. Estimating UK Regional Price Indices, 1974–96. *Regional Studies*, 2005, 3 (May), pp. 333–344.

HRONOVÁ, S., FISCHER, J., HINDLS, R., SIXTA, J. *Národní účetnictví: nástroj popisu globální ekonomiky* (National Accounting: a Tool for Description of the Global Economy). Prague: C. H. Beck, 2009.

CHLAD, M., KAHOUN, J. Factors Influencing the Rating of Regional Economic Performance or Reasons why Prague has Become the 6th Best Economically Performing Region of the EU. *Statistika, Economy and Statistics Journal*, 2011, 2 (June), pp. 4–23.

CZSO. *Evropský srovnávací program* (European Comparison Programme) [online]. Prague: Czech Statistical Office, 2012. [cit. 13.4.2012]. <http://www.czso.cz/csu/redakce.nsf/i/evropsky_srovnavaci_program>.

IBRD, WB. *Global Purchasing Power Parities and Real Expenditures – 2005 International Comparison Program*. Washington: International Bank for Reconstruction and Development, World Bank, 2008.

INTERNATIONAL COMPARISON PROGRAM. *ICP 2011* [online]. World Bank, 2009. [cit. 1.10.2012]. <http://go.worldbank.org/1XG3GMT780>.

JÍLEK, J., MORAVOVÁ, J. *Ekonomické a sociální indikátory. Od statistik k poznatkům* (Economic and Social Indicators. From Statistics to Knowledge). Prague: Futura, 2007.

ROOS, M. W. M. Regional Price Levels in Germany. *Applied Economics,* 2006, 13 (July), pp. 1553–1566.

STIGLITZ, J. E., SEN, A., FITOUSSI, J. P. et al. *Report by the Commission on the Measurement of Economic Performance and Social Progress* [online]. Commission on the Measurement of Economic Performance and Social Progress, 2009. [cit. 1.10.2012], <http://www.stiglitz-sen-fitoussi.fr/en/index.htm>.

ŠANDA, R. Sčítání založené na registrech – vytoužený cíl, nebo nutný kompromis? (The Census Based on Registers – Desired Goal or Necessary Compromise?). *Demografie,* 2012, 2 (June), pp. 162–165.

Foreign Direct Investment in Visegrad Four and the Main Trading Partners[1]

Ondřej Babuněk[2] | *University of West Bohemia, Pilsen, Czech Republic*

Abstract

This article is aimed at a comparison of the inflow of foreign direct investment per capita in the countries of the Visegrad four and their most important trading partners. For the comparison of countries there was exerted one statistical method - analyses of variance (ANOVA). It was found that there are statistical significant differences between the countries of the Visegrad four. It is shown that after the accession of Germany and Austria there is no statistically significant difference between all the countries. Germany and Austria are the geographically nearest countries of the Visegrad countries. After the inclusion of all trading partners who have ties to more countries, the Visegrad four is already evident that there are significant differences between countries. However, variances are statistically significant in the inflow of investment to all countries. In the work is shown that a country of the Visegrad four and its nearest trading partners are a compact unit in the inflow of foreign direct investment per capita.

Keywords

Foreign direct investment, population, ANOVA, the Visegrad four

JEL code

E22, E24, F21

INTRODUCTION

Foreign direct investment (FDI) has gained significant importance over the past decade as tool for accelerating growth and development of economies in transition. It is widely believed that advantages that FDI brings to the standard of living and prospects for economic growth of the host nation largely outweigh its disadvantages. International trade and FDI can be very effective ways of stimulating technological progress for a less-developed country. It is well known that there are many factors affecting FDI, such as intellectual property rights protection, economic stability and the political climate, labor market, opening policy, foreign exchange rate, relative wages and income convergence, financial and tax policy, GDP in the host country, bureaucratic corruption, an environmental policy and so on (Xu, 2008).

In recent years, FDI is worshipped in many places blindly, which has become an increasingly serious problem. Local government only focuses on the promoting function of FDI to the economic growth,

[1] This article is financially supported by the Student grant competition (Grant No. SGS-2012-036).

[2] Department of Economics and Quantitative Methods, Faculty of Economics, University of West Bohemia in Pilsen, Husova 11, 306 14 Pilsen, Czech Republic, e-mail: babunek@kem.zcu.cz, phone: (+420)774586751.

neglecting the negative effects of FDI, and does not understand the complex effect of FDI comprehensively. In the process of introducing FDI, the government pays attention to quantity rather than quality, and makes a clearance sale on natural resources, environment, market and even government tax in a competitive way. The bias of this policy and behaviour to attract foreign investment virtually increases the hidden trouble of immiserizing growth (Liu, 2011, p. 122).

For example, direct FDI effects in the Czech Republic are studied in many studies, which report that benefits are larger when investment comes in the form of FDI with direct foreign control rather than FDI in the form of a joint venture with a domestic company. In the Czech context, is argued that the estimated positive effects of FDI on performance are in some cases unrealistically high, and that the lack of suitable variables leads to an unsatisfactory estimation of self-selection. Using a data set of Hungarian firms, authors show that firms with foreign ownership outperform domestic firms (Hanousek, 2011).

However, such interventions may prove sub-optimal or even counterproductive. For instance, firms may have little choice but to engage in outward FDI if exporting from the home base is no promising alternative because of impediments such as distorted exchange rates and trade-related transaction costs. The optimal approach would then be to remove such distortions and, thereby, enable firms to reconsider their choice between exports and outward FDI. Restricting FDI-related offshoring to lower-cost locations such as China would involve serious trade-offs: while it might sustain domestic production and employment in the short run, such interference runs the risk of undermining the overall competitiveness of firms in the longer run. It should be noted in this context that there is little reason to blame FDI in China for hollowing-out Taiwanese manufacturing as the quantitative impact turns out to be rather small. Directing outward FDI to technologically leading host countries would have little effect unless the investing firms have sufficient absorptive capacity to make efficient use of superior technologies (Liu, 2011).

As the World Trade Organisation (WTO) only deals with "trade", the granting of incentives in the pre-production period and not for trade of goods creates problems in the measurement of adverse effects for other member states. By the time production and trade/exports have started, incentives given to attract investment have often ended (Oxelheim, 2008).

The paper is structured as follows. It begins with Section 1, which describes theoretical consequences. Section 2 includes methods and methodology including hypothesis. Section 3 provides results of research. Section 4 provides concluding remarks.

1 THEORETICAL CONSEQUENCIES

Inflow of foreign direct investments in particular countries varies from country to country, however object of this article are the states of the Visegrad four (V4) and the main trading partners of the V4. The V4 includes the Czech Republic, Hungary, Poland and Slovakia.

The correlation between exports and FDI suggests that government support of the mobile resources that are not tied to the UK, such as subsidies to R&D, management or technical training, designed to enhance the competitiveness, may affect also the competitiveness of its immobile resources, as measured by the export performance (Nachum, 2001). Employees are great mobile sources in some economies, which can´t be said for the Czech, who does not comply with these aspects. To support the mobility of employees are offered higher rewards.

The accruing inflow of FDI is affected preferential trade agreement such as the NAFTA, CEFTA, etc., concerning about an integration region. The main capital inflow of FDI to Korea after the liquidity crisis should take the form of cross-border M&A rather than Greenfield investment. It was caused merging firms and the main reason was scarcity of liquidity. In this consequence concerns increased

distinct cross-border M&A (Kim, 2009). Czech Republic is a member of many partnerships, mainly EU, OECD, V4 etc.

1.1 Trading partners of the V4

The main trading partners of the V4 are Russia, Ukraine, Italy, Great Britain, Germany, Austria and Romania. The main trading partner of the V4 is Germany, followed by Great Britain, France and Austria. These countries have the most linkages with countries of the V4. Russia, Ukraine and Romania have only one linkage with countries of the V4. These states were eliminated due to the only one linkage with countries of the V4 and this article is aimed to the remaining countries. These countries are Austria, Czech Republic, Germany, Hungary, Poland, Slovakia, France, Italy and Great Britain. The assessment of individual countries is referred in the following way:

The first areas are countries of the V4, then follow selected countries without France, Italy and Great Britain and at in the conclusions there are subsumed all the selected countries to illustrate the wider view of the analysis area. The reason for the second step was to determine whether the countries may differ, if there is included the distance between the individual countries, because the geographical distance between France, Italy and Great Britain is substantially higher than that between Germany and Austria, although on a global scale the distance between eliminated countries and the V4 including Germany and Austria is negligible.

1.2 Development of the population of the selected countries

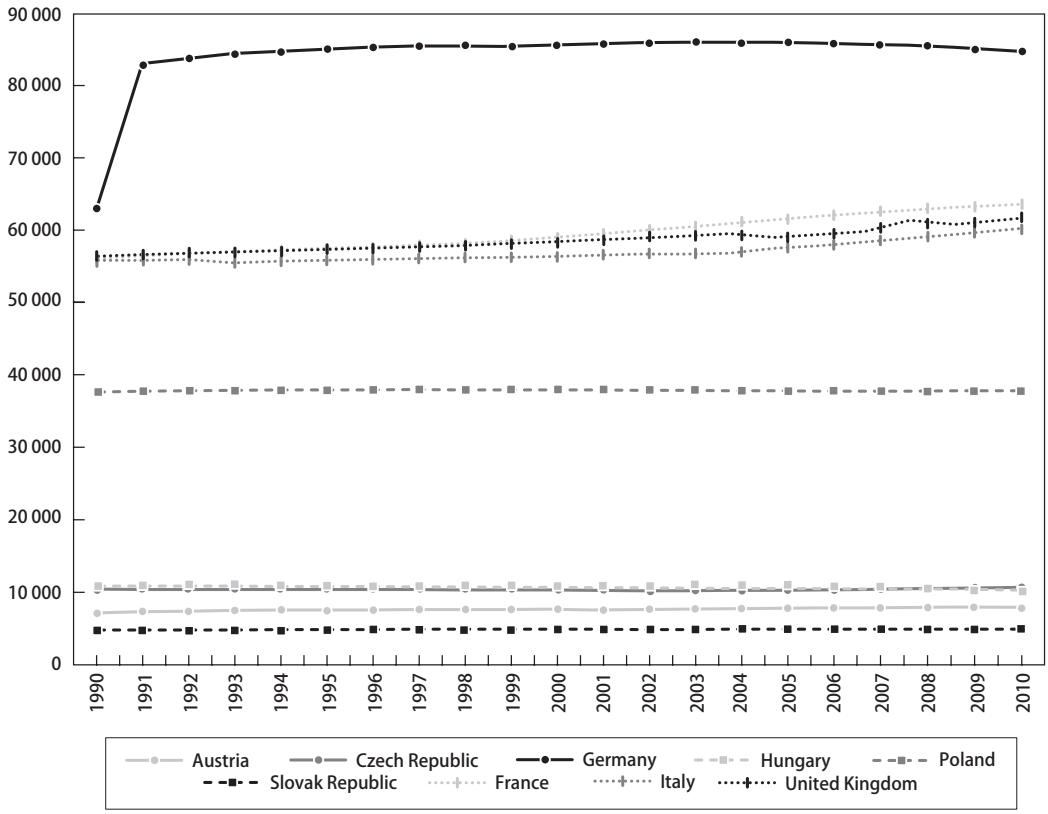
The market size also plays an important role on the pattern of FDI inflows. Countries with a larger market size appear to attract more FDI inflows (Mellahi, 2011). Market size is measured by the host country´s total population (Barassi, 2012). Milner (2006, pp. 205–206) argues that firms taking labor intensive stages of the production process to a lower cost location and transporting final and semi-processed products back to the home market. There are additional incentives to invest abroad where there are special exporting advantages (e.g. preferential market access terms), but it is predominantly production for a non-host country market.

Horstmann and Markusen, 1992 and Brainard, 1993 aimed to explain the high level of FDI between similar or even identical countries, and assume that the primary motivation for MNEs is to gain market access rather than to take advantage of differences in factor endowments. This stream of theory predicts that the host country's market size and trade cost would be vital in determining the level of FDI.

The development of the population of the selected countries is characterized by significant differences between countries. The absolute population size cannot be compared because Germany has 80 mil inhabitants. France, Italy and Great Britain have over 60 mil inhabitants, Poland has 38 mil, Czech Republic and Hungary have about 10 mil, Austria about 8 mil and Slovakia has roughly 5 mil inhabitants.

The following figure shows the development of population since 1993 to 2010. From the figure it is clearly perceptible that the development of population is constant during the period, nevertheless in Germany number of inhabitants between 1990 and 1991 grew. The cause of the population increase was the unification of the former West and East Germany. In the other countries population has grown since 1990, nevertheless Hungary is an exception, because population in this country was continuously falling down. In 1990 Hungary had 10.374 mil inhabitants but in 2010 only 10 mil inhabitants.

Figure 1 Population in the selected countries (in thousands USD)



Source: OECD

In the analysis there aren´t handled absolute numbers of inhabitants, but the relative ones, in order to get a comparison between countries with a higher explanatory value.

1.3 Development of inflow of foreign direct investment

Cuervo-Cazurra (2011) argued that firms choose states that are important physical and psychical distance and market attractiveness.

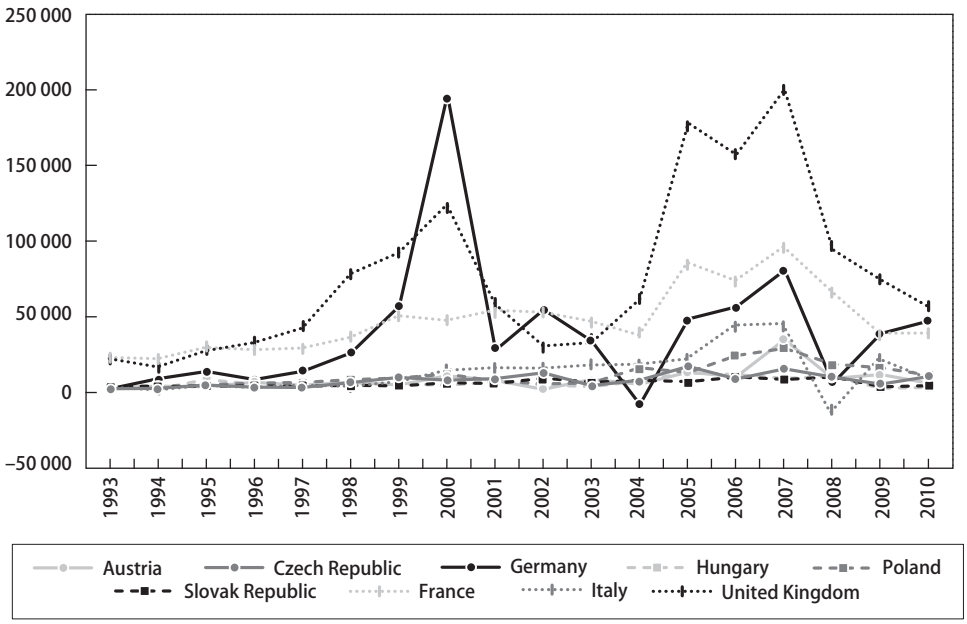
Altomonte (2003) argued that the Central and Eastern European countries (CEECs) display a great capacity in the attraction of FDI flows that this is likely due to the high degree of integration achieved among the CEECs: this structural characteristic of the Central and Eastern European region enhances the access to markets MNEs can serve from a location in the CEECs. From these states are generated increasing FDI inflows in the area.

Data from the Organisation for Economic Co-operation and Development (OECD) implies that the inflow of FDI is most pronounced in the United Kingdom (average is 65.634 billion USD per year), France (average is 39.521 billion USD per year) and Germany (average is 33.155 billion USD per year). To a lesser extent in Italy (average is 11.260 billion USD per year) and on the contrary in the remaining countries inflows of FDI are negligible in comparison with above mentioned countries. From the remaining selected countries has Poland the largest inflow of FDI (average is 7.537 billion USD per year).

Poland is followed Austria (average is 5.488 billion USD per year), Czech Republic (average is 4.821 billion USD per year), Hungary (average is 3.441 billion USD per year) and Slovakia (average is 1.732 billion USD per year).

The following figure shows development of inflow of FDI since 1993 and composes view on FDI in selected countries. The flow of investment is significantly different in individual countries and United Kingdom and Germany have the highest inflow of investment, by contrast, other countries have a significantly reduced inflow of investment.

Figure 2 Inflow of FDI into the selected countries (in USD)



Source: OECD

2 METHODS AND METHODOLOGY

The relevant data for the research was obtained from OECD statistics. For the testing of assumptions there must be used a methodology of consistent data acquisition, because obtained results wouldn´t have any value in different methodologies. Therefore, the data, which were published on web pages, year-books and official statistical records of the OECD have been selected. The same methodology to collect data is a pre-requisite for relevant results emanating from the data.

The main method is analysis of variance (ANOVA). This method serves as a tool for a comparison of the inflow of FDI per capita in selected countries. This analysis has many advantages. In the described method is mainly on the assessment of the impact of factors on the mean value of random variables, but its own analysis of the variance relates to the observed values, so it can talk about the analysis of variance, whose acronym was mentioned above. Analysis of variance can be distinguished by the number of influencing factors. For one character A is discussed about the analysis of variance with simple classifiers, in the case of two characters A and B on the analysis of variance for the dual classification, either with or without interaction. However, is significant whether these characters concur or not (Marek, 2007).

The analysis of variance with simple classifiers is based on model, which includes independent random variables with normal distribution N (μ, σ^2); μ, αi, σ^2 are unknown parameters. The hypoth-

esis that sign A has not an influence on observed random quantity X, correspond to the conjugate hypothesis $H_0 : \alpha_1 = \ldots = \alpha_i = 0$ with an alternative hypothesis H_1, that at least one α_i is different from the other α_k, i.e. that sign A has influence on observed random quantity X. For the testing a decomposition of sum of squares is used, where:

$$S_t = \sum_{i=1}^{I} \sum_{j=1}^{n_I} x_{ij}^2 - \frac{(x\,..)^2}{n}, \tag{1}$$

$$S_A = \sum_{i=1}^{I} \frac{(x_{i.})^2}{n_i} - \frac{(x\,..)^2}{n}, \tag{2}$$

$$\sum_{i=1}^{I} \sum_{j=1}^{n_I} (x_{ij} - \overline{x_{I.}})^2 = S_t - S_A . \tag{3}$$

Hypothesis H is tested at the significant level α with help of test criterion:

$$F = \frac{\dfrac{S_A}{I-1}}{\dfrac{S_e}{n-I}}. \tag{4}$$

After the rejection of the hypothesis at the chosen significant level there follows a testing of contrasts i.e. differences of the average values of pairs, if null (conjugate) hypothesis is rejected. Statistic is described below.

$$F = \frac{\dfrac{(x_{i.} - \overline{x_{k.}})^2}{I-1}}{\dfrac{S_e}{n-I}} \frac{n_i n_k}{n_i + n_k}. \tag{5}$$

Continue to testing the equality of groups variances i.e. Barttlets´ test. Statistic is described below.

$$B = \frac{1}{1 + \dfrac{1}{3(I-1)} \left(\sum_{i=1}^{I} \dfrac{1}{n_i - 1} - \dfrac{1}{n-I}\right)} \left[(n-I)\, In\ s^2 - \sum_{i=1}^{I} (n_i - 1)\, \mathrm{In}\ s_i^2\right]. \tag{6}$$

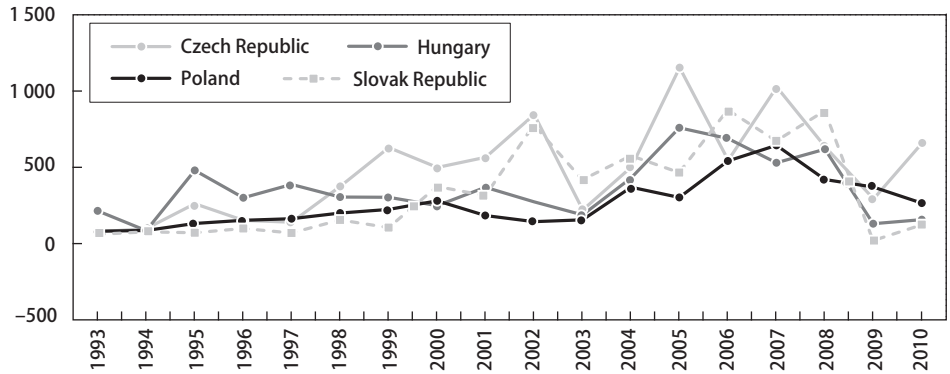
From the results of that analysis it is clearly evident which variable affects the observed random quantity X. Results of the analysis are shown in the following parts of the text (Karpíšek, 2007).

In this article there will be verified hypothesis H: Inflow of FDI per capita for V4 and their selected trading partners is the same in all countries.

3 RESULTS

This section describes the results of analyses, which were discussed in the previous section. The data was extracted from development of the inflow of FDI per capita in V4. From data obtained, that V4 have the similar development of inflow of FDI per capita. From 1993 to 2007 increased inflow of FDI per capita in all countries, but in 2008 and 2009 marked decline is declared. This decline is caused by difficulties in world economy, which proved after 2008 and are evident even in the following years, by contrast, in 2010 an investment growth is obvious.

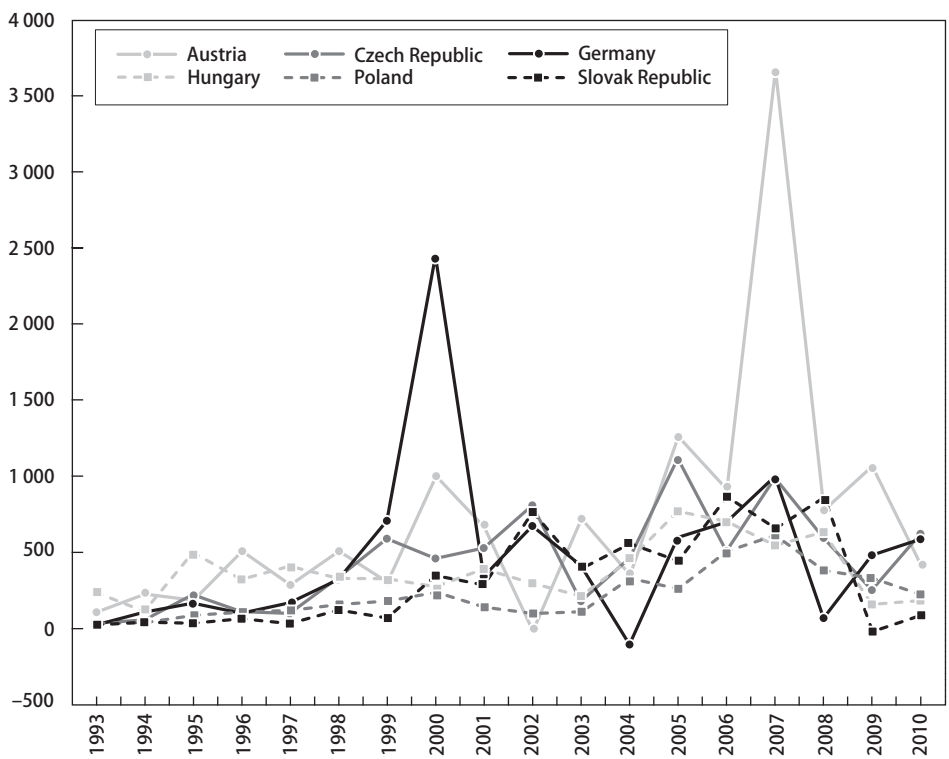
Figure 3 Inflow of FDI per capita into the V4 (in USD)



Source: OECD, own calculation

After the extension by Germany and Austria: The development of inflow of FDI per capita in those countries identical, however, Germany logged the largest fluctuation in 2000 and Austria in 2007. The value of these countries highly exceeded the other included countries. The development of inflow of FDI per capita has rising character, nevertheless in 2008 there was a significance decrease in flows to all countries.

Figure 4 Inflow of FDI per capita into the selected countries without France, Italy and UK (in USD)

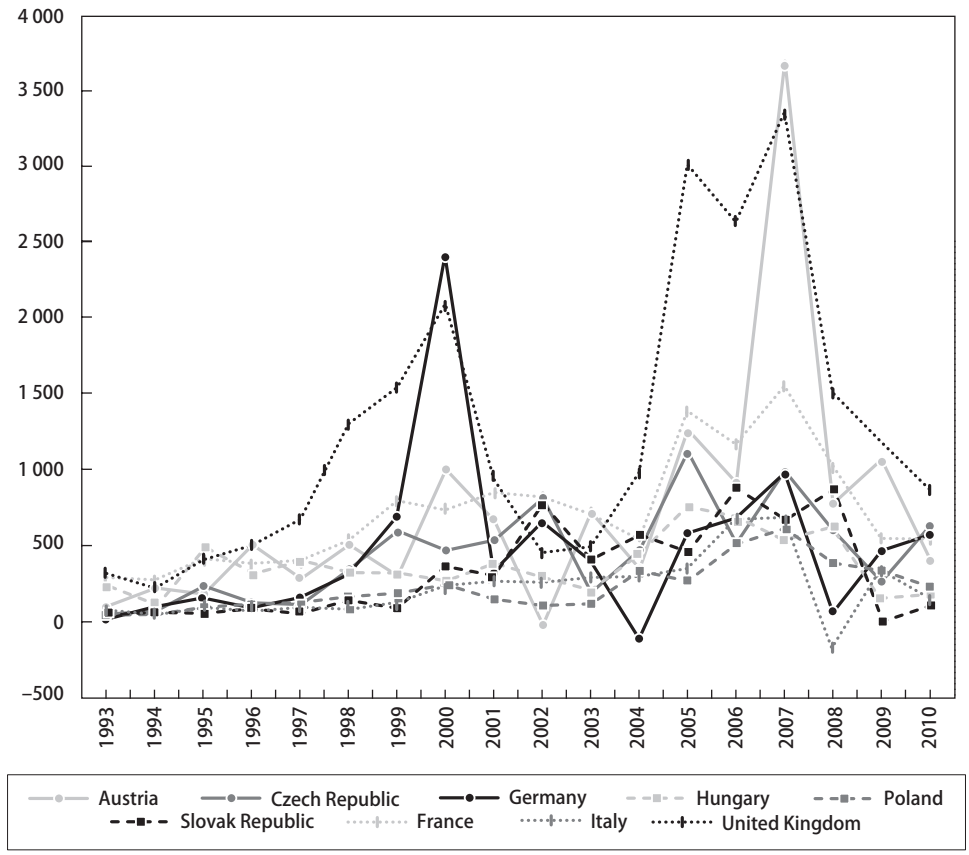


Source: OECD, own calculation

After the inclusion of all selected countries we see that investments substantially fluctuate and it is apparent that inflow of FDI per capita has two peaks. First peak was from 1999 to 2000. Those years point out that inflow of FDI per capita from previous years was significantly higher. Second peak was from 2005 to 2007. After these years came a fast and precipitate fall in 2008. However Germany and United Kingdom exceeded other countries in inflow of FDI per capita in 2000.

United Kingdom exceeds the remaining countries in the flow of investment from 2005 to 2007 and Austria exceeds the remaining countries in 2007.

Figure 5 Inflow of FDI per capita into the selected countries (in USD)



Source: OECD, own calculation

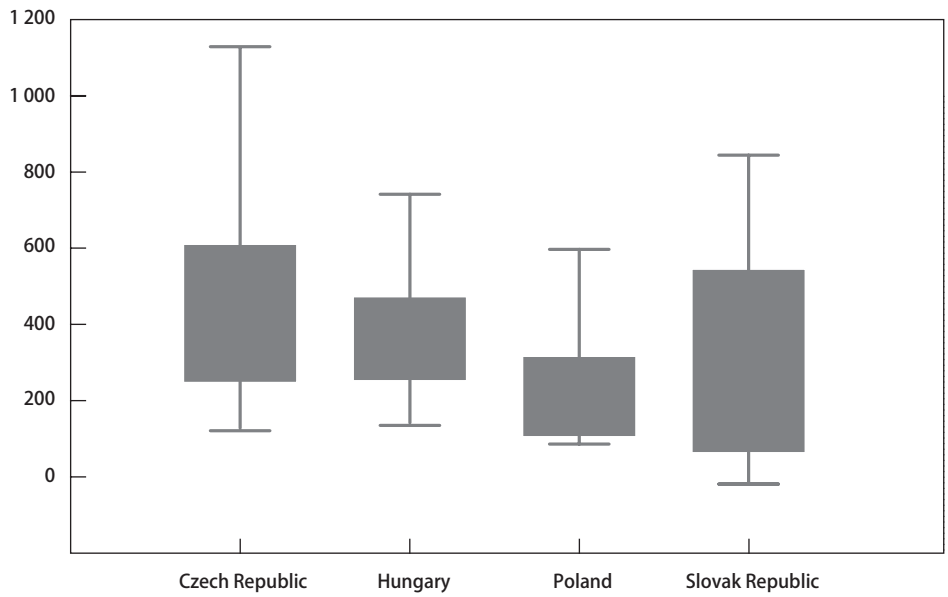
From the results of ANOVA it is evident, that the countries of V4 aren´t the same on α = 5% significant level (Table 1), but the value of statistics exceeds the critical value by only 0.12 and detailed results demonstrated that there is no statistically significant difference between individual countries which is opposite to the results of the ANOVA. Figure 6 points out this contention.

Table 1 ANOVA table for the V4

ANOVA α = 5%						
Source of variation	**SS**	**DF**	**MeanSq**	**Fratio**	**Pvalue**	**Fkrit**
Between selections	543 955.18	3	181 318.39	2.85	0.04	2.73
All selections	4 319 505.10	68	63 522.13			
Total	4 863 460.28	71				
ANOVA α = 1%						
Source of variation	**SS**	**DF**	**MeanSq**	**Fratio**	**Pvalue**	**Fkrit**
Between selections	543 955.18	3	181 318.39	2.85	0.04	4.08
All selections	4 319 505.10	68	63 522.13			
Total	4 863 460.28	71				

Source: Own construction

Figure 6 Box-and-Whisker Plot A (in USD per capita)



Source: OECD, own calculation

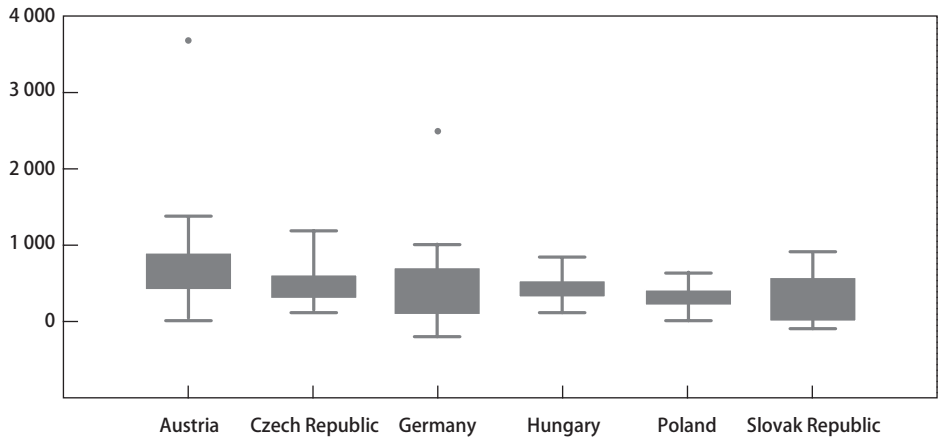
From the results of ANOVA is evident, that the countries of V4, Austria and Germany aren´t the same on α = 5% significant level (Table 2), but the value of statistics exceeds the critical value by only 0.57 and the detailed results demonstrated (Table 4) that there was found a statistically significant difference between Austria and Poland. Figure 7 points out this contention.

Table 2 ANOVA table for the selected countries without France, Italy and UK

ANOVA α = 5%						
Source of variation	**SS**	**DF**	**MeanSq**	**Fratio**	**Pvalue**	**Fkrit**
Between selections	3 028 914.81	5	605 782.96	2.87	0.017	2.30
All selections	21 476 329.14	102	210 552.24			
Total	24 505 243.95	107				
ANOVA α = 1%						
Source of variation	**SS**	**DF**	**MeanSq**	**Fratio**	**Pvalue**	**Fkrit**
Between selections	3 028 914.81	5	605 782.96	2.87	0.017	3.20
All selections	21 476 329.14	102	210 552.24			
Total	24 505 243.95	107				

Source: Own construction

Figure 7 Box-and-Whisker Plot B (in USD per capita)



Source: OECD, own calculation

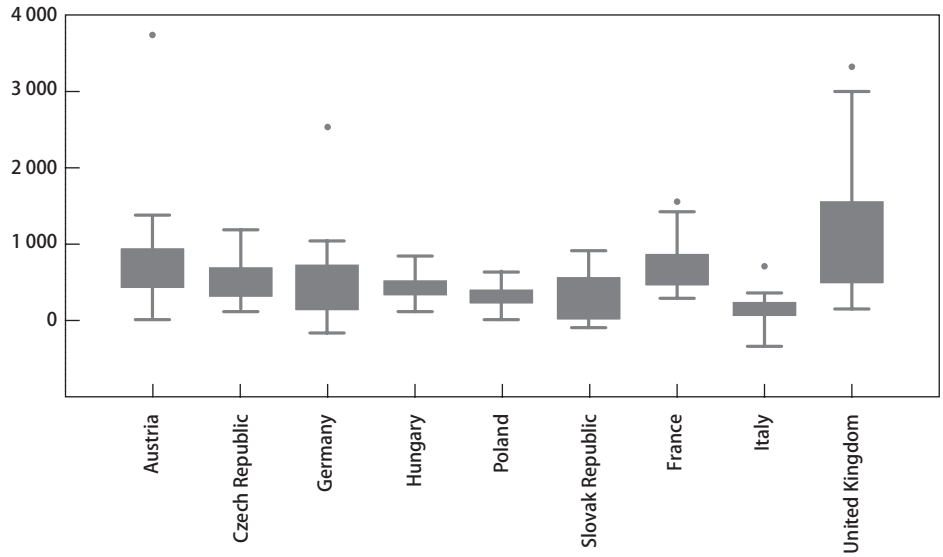
From the results of ANOVA is evident, that selected countries aren´t the same on α = 5% significance level (Table 3), and the statistics has substantially exceeded the critical value and that in the detailed analysis of relation between countries (Table 4) statistically significant differences were revealed between Czech Republic and United Kingdom, Germany and United Kingdom, Hungary and United Kingdom, Poland and United Kingdom, Slovak Republic and United Kingdom and Italy and United Kingdom. Figure 8 points out this contention.

Table 3 ANOVA table for the selected countries

ANOVA α = 5%						
Source of variation	**SS**	**DF**	**MeanSq**	**Fratio**	**Pvalue**	**Fkrit**
Between selections	15 021 642.20	8	1 877 705.27	7.16	4.86×10^{-8}	1.99
All selections	40 096 492.37	153	262068.57			
Total	55 118 134.57	161				
ANOVA α = 1%						
Source of variation	**SS**	**DF**	**MeanSq**	**Fratio**	**Pvalue**	**Fkrit**
Between selections	15 021 642.20	8	1 877 705.27	7.16	4.86×10^{-8}	2.62
All selections	40 096 492.37	153	262068.57			
Total	55 118 134.57	161				

Source: Own construction

Figure 8 Box-and-Whisker Plot C (in USD per capita)



Source: OECD, own calculation

Table 4 Significant differences between selected countries

Countries without Italy, France, UK	Test criterion	Critical value	Hypothesis
AU-PL	2.43769753	2.303493035	Reject
Selected countries	**Test criterion**	**Critical value**	**Hypothesis**
CZ-UK	2.446273772	1.999389853	Reject
GE-UK	2.453257862	1.999389853	Reject
HU-UK	3.069702354	1.999389853	Reject
PL-UK	4.243634936	1.999389853	Reject
SK-UK	3.489976782	1.999389853	Reject
IT-UK	4.363219568	1.999389853	Reject

Source: Own construction

For a detailed view results of variances of selected countries must still be given. From the analysis it can be argued that variances vary statistical significantly in all three investigated areas (for V4, for selected countries without France, Italy and United Kingdom and for all selected countries). The empirical results are shown in Table 5.

Table 5 Results of Bartlett´s test

Bartlett´s test	Test criterion	Critical value	Hypothesis
The Visegrad countries	11.43160439	7.815	Reject
Selected countries without Italy, France, UK	64.10817693	11.070	Reject
Selected countries	105.32065490	14.067	Reject

Source: Own construction

CONCLUSION

Foreign direct investment (FDI) has gained significant importance over the past decade as the tool for accelerating growth and development of economies in transition. It is widely believed that the advantages that FDI brings to the standard of living and prospects for economic growth of the host nation largely outweigh its disadvantages. International trade and FDI can be very effective ways of stimulating technological for a less-developed country. It is well known that there are many factors affecting FDI, such as intellectual property rights protection, economic stability and the political climate, labor market, opening policy, foreign exchange rate, relative wages and income convergence financial and tax policy, GDP in the host country, bureaucratic corruption and environmental policy and so on (Xu, 2008).

This article has compared countries of the V4 and their main trading partners. The object of comparison was to find out whether the inflow of FDI per capita in countries of V4 is similar and therefore it can be argued that inflow of FDI per capita to these countries is identical for all countries.

Further emerged that the V4 countries, Germany and Austria have a similar inflow of FDI per capita. Difference between Austria and Poland was only one distinction in this section, however, this distinction wasn´t large. Germany and Austria are the closest geographical neighbours of selected countries of V4. For future studies it will be centre of many researches. This factor may influence inflows of FDI into the selected countries.

In comparison of all selected countries including France, Italy and United Kingdom it was found that distinction between selected countries is statistically significant. The most important element which significantly affected results of all selected countries was the United Kingdom. United Kingdom significantly exceeds inflow of FDI per capita compared with Czech Republic, Germany, Hungary, Poland, Slovakia and Italy. To illustrate the analysis results of the analysis have been displayed in a Box-and-Whisker Plot. These results confirmed conclusions mentioned above.

Variability of inflow of FDI per capita is considerable in all selected countries and all countries show a marked difference. Variability is dependent on a number of factors for example economic growth, GDP, unemployment, competition etc. This factors weren´t subsumed in this article.

ACKNOWLEDGEMENTS

A special thanks goes to Eva Babuňková for her help with this paper.

References

ALTOMONTE, C., GUAGLIANO, C. Comparative Study of FDI in Central and Eastern Europe and the Mediterranean. *Economic Systems,* 2003, No. 2, pp. 223–246.

BARASSI, M. R., ZHOU, Y. The Effect of Corruption on FDI: A Parametric and Non-Parametric Analysis. *European Journal of Political Economy,* 2012, No. 3, pp. 302–312.

BRAINARD, S. L. A Simple Theory of Multinational Corporations and Trade with a Trade off Between Proximity and Concentration. *NBER Working Paper,* 1993, No. 4269, pp. 1–48.

CUERVO-CAZURRA, A. Selecting the Country in which to Start Internationalization: The Non-Sequential Internationalization Model. *Journal of World Business,* 2011, No. 4, pp. 426–437.

HANOUSEK, J. KOČENDA, E., MAUREL, M. Direct and Indirect Effects of FDI in Emerging European Markets: A Survey and Meta-Analysis. *Economic Systems,* 2011, No. 3, pp. 301–322.

HORSTMANN, I. J., MARKUSEN, J. R. Endogenous Market Structures in International Trade (Natura Facit Saltum). *Journal of International Economics,* 1992, No. 1–2, pp. 109–129.

KARPÍŠEK, Z. *Kvantitativní metody v řízení* (Quantitative Methods in Management). Brno: Akademie Sting, 2007.

KIM, Y. Cross-Border M&A vs. Greenfield FDI: Economic Integration and its Welfare Impact. *Journal of Policy Modeling,* 2009, No. 1, pp. 87–101.

LIU, W., NUNNENKAMP, P. Domestic Repercussions of Different Types of FDI: Firm-Level Evidence for Taiwanese Manufacturing. *World Development,* 2011, No. 5, pp. 808–823.

MAREK, L. *Statistika pro ekonomy: aplikace* (Statistics for Economists: Application). Prague: Professional Publishing, 2007. ISBN 978-80-86946-40-5.

MELLAHI, K., DEMIRBAG, M., RIDDLE, L. Multinationals in the Middle East: Challenges and Opportunities. *Journal of World Business,* 2011, No. 4, pp. 406–410.
MILNER, C., REED. G., TALERNGSRI, P. Vertical Linkages and Agglomeration Effects in Japanese FDI in Thailand. *Journal of the Japanese and International Economies,* 2006, No. 2, pp. 193–208.
NACHUM, L., DUNNING, J. H. The International Competitiveness of the UK and its Multinational Enterprises. *Structural Change and Economic Dynamics,* 2001, No. 3, p. 277–294.
OXELHEIM, L., GHAURI, P. EU-China and the Non-Transparent Race for Inward FDI. *Journal of Asian Economics,* 2008, No. 4, pp. 358–370.
XU, W., HU, D., LEI, A., SHEN, H. FDI Chaos and Control in China. *Research in International Business and Finance,* 2008, No. 1, pp. 17–28.

Stastistical Survey of Non-Formal Education

Ondřej Nývlt[1] | *University of Economics, Prague, Czech Republic*

Abstract

At present it is not possible to consider the education system only from a point of view of formal education focused on a programme within a regular education system. Labour market flexibility and new requirements on employees create a new domain of education called non-formal education. Is there a reliable statistical source with a good methodological definition for the Czech Republic? Labour Force Survey (LFS) has been the basic statistical source for time comparison of non-formal education for the last ten years. Furthermore, a special Adult Education Survey (AES) in 2011 was focused on individual components of non-formal education in a detailed way. In general, the goal of the EU is to use data from both internationally comparable surveys for analyses of the particular fields of lifelong learning in the way, that annual LFS data could be enlarged by detailed information from AES in five years periods. This article describes reliability of statistical data about non-formal education. This analysis is usually connected with sampling and non-sampling errors.

Keywords

Labour force survey, non-formal education, sampling errors, non-sampling errors

JEL code

I2, J2

INTRODUCTION

Higher or lower education is important for explanation of fundamental social, economic or demographic differences. For example, it is the most important factor for analysis of the labour market. If divided by educational attainment unemployment rate in the Czech Republic according to the ILO definition (in the age group 15–64 years) in 1st quarter of 2011, university graduates permanently show a low unemployment rate (2.5%) and similarly persons who have secondary education with an A-level examination (5.2%). A high unemployment rate grows in the group of persons with basic education (28.4%), and an above average unemployment rate also pertains to the large group of those who have got secondary education without an A-level examination, including persons with apprenticeship certificates (8.2%) (News Releases Employment and Unemployment, 2011).

Significant economic changes bring new requirements on the labour market and force employees to gain new skills. The labour market creates demand for education but does not directly focus on formal educational system. The demand is oriented rather on a particular type of gaining skills. The non-formal education plays an important role in this case. For this reason, the basic indicator of education (life-long learning indicator) used for the international comparison works on the assumption of participation in formal or non-formal education.

[1] University of Economics, Prague, nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic; Czech Statistical Office, Na padesátém 81, 100 82 Prague, Czech Republic, e-mail: ondrej.nyvlt@czso.cz.

Non-formal education covers all taught organised learning activities outside the regular education system. A non-formal learning activity is defined as being organised like a course, conference or seminar for which the interviewee has applied and has participated in. It could be for a short or a longer period with possible minor breaks (EU Labour Force Survey, 2011).

1 METHODOLOGY

Sample surveys are usually connected with sampling and non-sampling errors. The latter are a result, for instance, of administrative drop-outs of dwellings out of the sample, intentional non-response or errors produced by filling in the questionnaire (Employment and Unemployment in the Czech Republic, 2011). With these errors, one cannot determine the deviation of estimate without a rather wide knowledge of the population. On the other hand, the sampling errors which arise by applying characteristics of the sample to the population can be interpreted by means of confidence intervals. The confidence intervals are intervals built around the point estimate in such a way that there is a certain probability that the value of an estimated characteristic is just within this interval. The most widely used is a 95% confidence interval, i.e. an interval within which the actual value of the estimated characteristic is found with 95% probability.

From another point of view it is possible to divide errors into observational and non-observational ones. Observational errors are deviations of the answers of respondents from their true values from the measure; for our purposes, these are measurement errors. Observational errors are conveniently categorized according to different sources – the interviewer, the respondent, the questionnaire, and the mode of data collection. Interviewer's errors are associated with effects on respondents' answers stemming from different ways in which interviewers administer the same survey (Measurement Errors in Surveys, 1991). But once we begin to consider the possibility of alternative wordings, we seem to presume that there is something of which they are alternative versions: an abstract question that we can ask in different ways (The Psychology of Survey Response, 2000). The respondent error shows that different respondents have been found to provide data with different amount of error, because of different cognitive abilities or differential motivation to answer the questions well (Measurement Errors in Surveys, 1991). The meaning that we get from a word or a sentence must be relatively stable across people (The Psychology of Survey Response, 2000). The interpretation of a sentence has to be at least somewhat immune to differences in the amount of knowledge about the concepts.

Non-observational errors are those arising because of measurement not taken on part of the population and there are viewed as arising from three sources – coverage, non-response and samplings (Measurement Errors in Surveys, 1991). The sample in the Labour Force Survey is based on the evidence of dwellings in the area of the Czech Republic where only registered flats are involved. The problem is that collective households (e.g. lodging house, old people's home) have a totally different demographic and social structure than registered flats (higher number of pensioners and foreigners), so the coverage has a great impact on the basic results of LFS. According to the statistical law, the citizens of the Czech Republic can refuse to participate in household surveys and the non-response significantly influences the quality of LFS data. This description shows that sampling error is only a part of the measurement errors in survey sample.

It is important to determine the time period. However, many surveys items depend on both the interview time and event time. These items include the ubiquitous questions about whether an event happened in the last few weeks, months, or some other time interval. It is necessary to search in memory for the event information and to locate this information with respect to the bounds of the reference period (The Psychology of Survey Response, 2000).

2 DATA

The Adult Education Survey (AES) should take place every 5 years starting from 2011 and is designed to give detailed information on the participation of individuals in education and training activities.

A pilot survey was held in 2007. The reference period for the participation in education and training activities is twelve months prior to the interview. In the Czech Republic the collection of data was implemented in the second half of 2011 on the sample of 9 500 households. Interviewers asked all persons in households at the age of 18–69 years.

The basic source for providing data about non-formal education is now the Labour Force Survey. The LFS is a continuous survey whose results are assessed on quarterly basis. Quarterly, the sample comprised almost 25 thousand dwellings in the area of the Czech Republic (0.6% of all dwellings permanently lived in), including more than 50 thousand respondents aged 15 and over. This sample size allows making estimates of labour market indicators as well as the educational indicator (formal or non-formal education). Weight adjustments have been made to all the sample data according to the age, gender and region structure of the population based on mid-quarter population projection for the Labour Force Sample Survey (Employment and Unemployment in the Czech Republic, 2011).

3 DESCRIPTION OF THE PILOT SURVEY OF NON-FORMAL EDUCATION

Labour Force Survey is a relatively time demanding survey and is primarily focused on data on economic activity or inactivity. The question about non-formal education is at the end of the questionnaire. In addition to that, the determination of non-formal education is a very complicated methodological process. If we express such a complicated process only by one question, it can lead to a relatively high underestimation. In this case it is necessary to describe the non-formal education in a more detailed way.

The pilot survey was conducted on a sample of the standard LFS module in all dwellings in the 4th quarter of 2011. For the 4th quarter of 2011, three basic questions about participation in non-formal education were implemented. Before the 4th quarter of 2011, the information on variable participation in non-formal education was collected only from one question.

Table 1 Question on reference period in non-formal education (standard LFS)

Filter	Name of question	Question and answers	Code
15–69	VzdNform	Did s/he attend education within non-formal education (in last 4 weeks)?	
		Yes	1
		No	2

Source: Czech Republic, LFS

In the 4th quarter of 2011 the first question was aimed at any courses, the second on any seminars or workshops and the third on private lessons outside the formal education system. When the person at least once answers the question Yes, this person is afterwards considered as person who attended non-formal education. The questions are associated with a reference period of fourth weeks, both in the pilot survey and standard questionnaire of LFS.

Table 2 Questions about participation in non-formal education (pilot survey, 4th quarter 2011)

Filter	Name of question	Question and answers	Code
15+	H0001 (NFECOURSE)	Within the last **4 weeks** did s/he attend any courses?	
		Yes	1
		No	2
15+	H0003 (NFEWORKSHOP)	Within the last **4 weeks** did s/he attend any seminars or workshops?	
		Yes	1
		No	2
15+	H0005 (NFELESSON)	Within the last **4 weeks** did s/he receive private lessons outside the formal education system?	
		Yes	1
		No	2

Source: Czech Republic, LFS

The important factor of a survey is to define the reference period. In case of non-formal education it is important because the respondent sometimes forgot during the time, that he has attended the non-formal education. Within the framework of the pilot survey one question about attending education within non-formal education in last 12 months was added. This reference period allows the comparison with reference period in AES.

Table 3 Questions about reference period in non-formal education (pilot survey, 4th quarter 2011)

Filter	Name of question	Question and answers	Code
H0001=2 and H0003=2 and H0005=2	H0009	Did s/he attend education within non-formal education (**in last 12 months**)?	
		Yes	1
		No	2

Source: Czech Republic, LFS

4 MAIN RESULTS

Generally, it is necessary to define the framework of a sampling and non-sampling error in a survey. The sampling error can be interpreted by means of 95% confidence intervals for the estimate.

The confidence intervals are calculated for the sample size in a given quarter. In order to calculate confidence intervals of aggregates the following formula should be used for the basic aggregate (Employment and Unemployment in the Czech Republic, 2011).

$$95\%\ C.I.\ of\ estimate\ Y = y \mp 1.96 \times s_y,\ where\ s_y \cong N \times \sqrt{(1-f) \cdot \frac{\frac{y}{N} \times (1 - \frac{y}{N})}{f \times N}}, \quad (1)$$

where: N is the size of the population,
y is the estimate of aggregate Y in the population,
f is the respective proportion of sample, $f = n / N$.

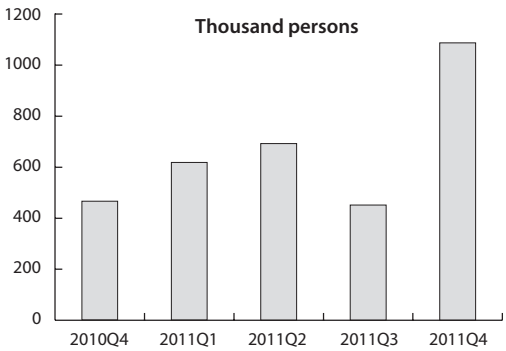
When substituting the variables into the above-mentioned formula we get the confidence intervals of number of participants in non-formal education. The resulting 95% confidence interval for the estimate of the number of persons attending non-formal education in 2nd quarter of 2011 is approx. 684.4 +/– 20.9 thousand, i.e. there is a 95% probability that the actual number of participants in non-formal education in the Czech Republic was not below 663.5 thousand and not above 705.3 thousand. The resulting 95% confidence interval for the estimate of the number of participants in non-formal education in 4th quarter of 2011 is approx. 1 127.6 +/– 26.0 thousand, i.e. there is a 95% probability that the actual number of participants in non-formal education in the Czech Republic was not below 1 101.6 thousand and not above 1 153.6 thousand.

It is a fact that sampling error itself cannot explain differences between the 2nd quarter of 2011 and the 4th quarter of 2011. It is clear that asking three separate questions about non-formal education instead of one has a significant impact on data and time-series. Such change in methodology influences the respondent's ability to better recall previous education activities. Therefore, the number of non-formal education participants in the 4th quarter of 2011 more than doubled in comparison with the 2nd quarter of 2011.

[2] A course is defined as "a planned series of single learning activities in a particular range of subject-matters offered by a provider". Courses are typically subject oriented and they are taught by one or more persons specialised in the field(s) of education and training. They may take place in one or more settings/environments.

[3] Sessions combining theoretical instruction with "hands-on" training provided during a conference or congress.

Figure 1 Non-formal education, population 15–69, Czech Republic



Source: Czech Republic, LFS

5 REFERENCE PERIOD

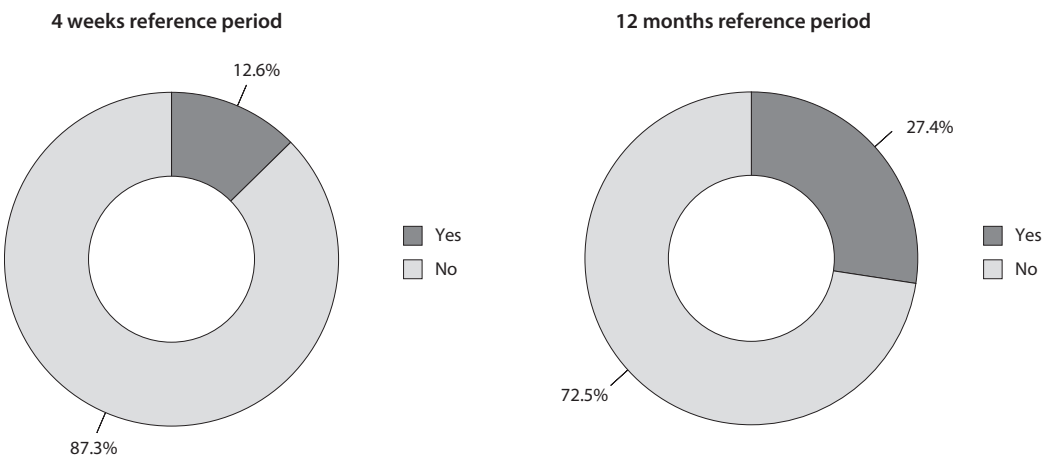
The reference period in standard LFS (the last 4 weeks) is defined differently than in AES (the last 12 months). From this perspective, the impact on the total number of participants in non-formal education was, therefore, determined.

On average, more than 1 out of 10 respondents participated in non-formal education in the last 4 weeks. Considering the last 12 months, it was more than 1 out of 4. The effect of a different reference period is clear and obvious.

6 DETAILED RESULTS

The main results show underestimation of non-formal education from a standard questionnaire

Figure 2 Participation in non-formal education, reference period comparison



Source: Czech Republic, LFS

in the Labour Force Survey. From this result, we cannot obtain a detailed explanation of this underestimation. But it is clear that the underestimation is created mainly by non-sampling errors. A detailed analysis according to the regions of the Czech Republic shows the basic differences. During the year, there is not a preference of a specific mode of questioning and therefore the model does not count with the mode effect. The results express the bias from the part of the interviewer, respondent or questionnaire errors. For analytic purposes the comparison of the 2nd quarter and the 4th quarter of 2011 was used. The 3rd quarter generally is seasonally influenced by the summer holiday. In this quarter the number of participants within non-formal education is lower.

The interregional comparison shows the divergence in results. In some regions the differences were high, in others insignificant. Generally, in the Czech Republic in the comparison between the 2nd and 4th quarter of 2011, there was a double difference. In the 2nd quarter of 2011, of 9 013.0 thousand persons aged 15 or more 684.4 thousand attended the non-formal education, in the 4th quarter of 2011

of 9 013.5 thousand persons aged 15 or more 1 127.6 thousand attended the non-formal education. Relatively, in the 2nd quarter of 2011 the proportion of participants in non-formal education was 7.6%, in the 4th quarter of 2011 already 12.5%. We can compare relative values between the 2nd quarter of 2011 and the 4th quarter of 2011 according to regions (NUTS3). We express the comparison by index in percent. The high value of this index was in Jihomoravský region, (where the index was 298.8%), in Praha (212.1%) and in Středočeský region (198.0 %). On the contrary, the lowest value of index was in Karlovarský region (100.9%), then in Zlínský region (109.4%) and region Vysočina (117.8%). Differences between regions were huge and they indicate the problem of quantification of non-formal education in the Czech Labour Force Survey. The comparison between the regions of the Czech Republic was used because the organizational structure of regions is relatively independent of the central level. It is obvious from the results that each region approaches this problem differently. Different approaches influence the final results for the Czech Republic.

Table 4 Participation in non-formal education by region in the Czech Republic (2nd quarter 2011, 4th quarter 2011)

Region	2011q2 (1 question)			2011q4 (3 questions)		
	Total	Participation in non-formal education		Total	Participation in non-formal education	
		Absolutely (thousand)	Relatively (in %)		Absolutely (thousand)	Relatively (in %)
Czech Republic	9 013.0	684.4	7.6	9 013.5	1 127.6	12.5
Praha	1 097.4	100.9	9.2	1 099.9	2 14.5	19.5
Středočeský	1 073.0	79.4	7.4	1 079.5	158,1	14.7
Jihočeský	545.6	36.6	6.7	545.2	60.2	11.0
Plzeňský	491.2	51.8	10.5	491.4	65.4	13.3
Karlovarský	262.0	20.7	7.9	261.5	20.9	8.0
Ústecký	706.8	62.7	8.9	705.8	87.0	12.3
Liberecký	373.4	27.8	7.4	373.3	43.3	11.6
Královéhradecký	473.6	40.7	8.6	473.0	66,8	14.1
Pardubický	440.9	39.3	8.9	440.7	55.6	12.6
Vysočina	439.4	36.3	8.3	438.8	42.7	9.7
Jihomoravský	991.4	31.1	3.1	990.9	92.9	9.4
Olomoucký	548.9	30.4	5.5	547.8	44.8	8.2
Zlínský	506.8	40.9	8.1	505.7	44.7	8.8
Moravskoslezský	1 062.6	85.9	8.1	1 059.9	130.8	12.3

Source: Czech Republic, LFS

These results indicate that the original model of the questionnaire (before 4th quarter of 2012) was inconvenient due to its vagueness. The original model did not specify the individual forms of non-formal education. The total length of the questionnaire of the Czech Labour Force Survey causes generally simplifying of questions which are methodologically very difficult. Furthermore, the situation with a relatively long questionnaire can often lead to satisfaction with a simple answer No in the question about non-formal education. When the question about non-formal education is asked in an exhaustive form it is possible that the original answer No will be changed to Yes. In this case we can talk about a combination of an interviewer and respondent error. This result is documented in Table 5. In this table 95% confidence intervals to estimates of non-formal education for individual regions are calculated. It is necessary to say that in case of regions (NUTS3) the estimates of non-formal education are relatively low. For this reason the sampling errors are relatively higher than in case of the whole Czech Republic. In five regions are relative 95% confidence intervals to estimates of non-formal education higher than 10%, for the Czech Republic it is only 2.3%. In spite of the exception of two regions (Karlovarský and Zlínský region), there is an absolute and relative difference between the 2nd quarter of 2011 (one question) and the 4th quarter of 2011 (three questions) higher than 95% confidence intervals to es-

timates of non-formal education. For example in Olomoucký region the difference between the 2nd and 4th quarter of 2011 was 14.4 thousands persons (47.6% growth), 95% confidence intervals to estimates of non-formal education was only 5.6 thousands (relatively 12.6%). From these results it is obvious that non-sampling error plays an important role in the interpretation of the results of non-formal education from the Labour Force Survey data.

Table 5 Differences and 95% confidence intevals to estimate the number of participations in non-formal education in the Czech Republic (in thousands, in %)

Region (NUTS3)	Growth (2011q4, 2011q2) – absolutely	Growth (2011q4, 2011q2) – relatively	95% confidence interval (abs. +/–) 2011q4	95% confidence interval (rel. +/–) 2011q4
Czech Republic	443.2	64.7%	26.0	2.3%
Praha	113.6	112.1%	14.3	6.8%
Středočeský	78.8	98.0%	9.4	6.0%
Jihočeský	23.5	64.4%	5.3	8.8%
Plzeňský	13.6	26.2%	5.5	8.5%
Karlovarský	0.1	0.9%	2.9	14.1%
Ústecký	24.3	39.0%	8.2	9.5%
Liberecký	15.5	55.9%	4.8	11.1%
Královéhradecký	26.1	64.2%	6.4	9.6%
Pardubický	16.3	41.6%	5.3	9.6%
Vysočina	6.4	17.8%	4.6	11.0%
Jihomoravský	61.8	198.8%	7.9	8.4%
Olomoucký	14.4	47.6%	5.6	12.6%
Zlínský	3.7	9.4%	5.1	11.3%
Moravskoslezský	44.9	52.7%	9.0	7.0%

Source: Czech Republic, LFS

CONCLUSION

The pilot project provided much more detailed information on participation in non-formal education than standard core LFS. It proved the fact that a change in methodology leads to different, better quality results. The pilot survey indicated that the indicator of non-formal education is significantly underestimated. This error can be due to the interviewer, respondent or methodological side. The interviewers often do not gain relevant information on participation in non-formal education. Furthermore, the respondents react negatively to the survey and reduce their answers. Finally we can say that sampling error does not play as important role as a non-sampling error, and this conclusion is valid even in the regions (NUTS3) of the Czech Republic. The pilot survey indicated a possible solution to the problem. Based on the output of the pilot survey, the wording was changed for the variable expressing non-formal education as follows:

Before the pilot (2011Q3):
Did s/he attend education within non-formal education (in last 4 weeks)?

After the pilot (2012Q1):
Did s/he attend course, seminar, workshop or private lesson within the last 4 weeks?

References

Analysis of Survey Data. Southampton: University of Southampton, 2004.
CZSO. *Employment and Unemployment in the Czech Republic as Measured by the Labour Force Sample Survey* [online]. Prague: Czech Statistical Office, 2011. <http://www.czso.cz/eng/edicniplan.nsf/aktual/ep-3?opendocument#31>.
CZSO. *News Releases Employment and Unemployment* [online]. Prague: Czech Statistical Office, 2011. <http://www.czso.cz/csu/csu.nsf/engkalendar/2011-zam>.
EUROSTAT. *EU Adult Education Survey. Explanatory Notes.* Brussels: Eurostat, European Commission, 2012.
EUROSTAT. *EU Labour Force Survey. Explanatory Notes.* Brussels: Eurostat, European Commission, 2011.
Measurement Errors in Surveys. New Jersey: John Wiley & Sons. 2004.
The Psychology of Survey Response. Cambridge: Cambridge University Press, 2005.
THOMPSON, S. K. *Sampling.* New York: Pennsylvania State University, 2002.

Income Expectations of University Students: Sample of Selected Economic Universities in the Czech Republic, England and Poland

Kateřina Maršíková[1] | *Technical University of Liberec, Czech Republic*
Aleš Kocourek[2] | *Technical University of Liberec, Czech Republic*

Abstract

The objective of the paper is to analyze students' perceptions and expectations of their future incomes. By doing so, authors bring a possible quantitative argument about the rate of return of investment into higher education based on data from selected economic universities, and thus open space for discussion about financing of Czech public universities from public or private sources. The authors used data from a large survey among students of selected faculties of economics and compared the results from Czech, Polish and English respondents. All the performed tests have indicated and identified one critical finding that confirmed results of previous surveys: the value of the spot expected rates of return on investment to tertiary education for the Czech Republic never falls below 8.45 per cent, even when allowed for gender of respondents, knowledge about income of respondents' friends, or level of education of respondents' parents.

Keywords

Tertiary education, rate of return, income, tuition fee, investment

JEL code

I22, H52

INTRODUCTION

The economic situation in many countries all around the world is forcing their governments to consider some alternative (private) sources of financing for ever more important and also more demanding public universities. One of the possibilities on side is tuition fees. This article provides new facts and figures casting some light on the willingness or readiness of the students themselves to pay for their tertiary education. It compares the data from three different countries and introduces a couple of factors ("agents")

[1] Technical University of Liberec, Faculty of Economics, Studentská 2, 461 17 Liberec 1, Czech Republic. Corresponding author: e-mail: katerina.marsikova@tul.cz, phone: (+420)485352344.

[2] Technical University of Liberec, Faculty of Economics, Studentská 2, 461 17 Liberec 1, Czech Republic. Corresponding author: e-mail: ales.kocourek@tul.cz, phone: (+420)485352424.

influencing students' income expectations. These are used due to a lack of data about real earnings classified according to the field of study in each county.

Technological advance has driven up the demand for skills; human capital is nowadays even more important determinant of economic competitiveness than in the past – a core argument underpinning increased education spending in the USA, the crisis notwithstanding. To compete internationally, countries need mass high-quality systems of tertiary education. But public budgets face competing imperatives such as population ageing and increased pressures on medical spending. Countries typically pursue three efficiency goals in higher education: larger quantity, higher quality, and constant or falling public spending. Systems that rely on public finance can generally achieve any two of the goals, but only at the expense of the third one: a system can be large and tax-financed, but with doubts about quality (France, Germany, Greece, Italy); or high-quality and tax-financed, but small (the UK until 1990); or large and high-quality, but fiscally expensive (as in Scandinavia). There is nothing illogical about the last option, but it has been unsustainable in most countries. The only realistic route – by which tertiary education could avoid being starved of funds – has been or could be (in some countries as the Czech Republic) supplementing the public spending on a significant scale with private finance (Barr, 1993).

According to Barr, tertiary education creates benefits beyond those to the individual – social benefits in terms of growth, the transmission of values, and the development of knowledge for its own sake. All these justify continuing taxpayer support. However, graduates typically also receive private benefits – higher earnings, more satisfying jobs, greater enjoyment of leisure – making it efficient and fair they cover a part of the costs. However, they should bear these costs when they can afford it, when they receive the private benefits of their university degrees, i.e. as graduates, not already as students (Barr 1993, 2010).

According to the theory of human capital, the choice of level of education, its length and field of study depends on returns to this investment (Becker, 1993). Also people's choice of an educational path is based on what they see as the optimum financial return derived by them from such a choice (Wolter, Weber, 1999). Barr (2010) points out a set of four objectives for tertiary education. Policy should seek to:

- Widen participation, both for equity reasons and on efficiency grounds, since any country cannot afford to waste a talent.
- Strengthen the quality of teaching and research.
- Protect the autonomy of universities, which is desirable both for its own sake and, more instrumentally, because autonomy and quality are strongly linked.
- Protect the fiscal background.

1 RESULTS OF THE SURVEY

The idea is rather fundamental: a rational student is not willing to pay for his/her university studies more than how much the additional value – the degree brings to him / her – is. The results of research made in previous years show that students perceive the investment into the higher education as something very expediential (Urbánek et al., 2009).

To find out students expectations, a questionnaire survey has been conducted to collect the data from the first year students at different universities, but all at selected faculties of economics, representing both Prague and regional ones. Students were asked about their expected income after graduation and after ten years of working experience in both cases – with a high-school degree only and with a master degree. They also provided the information about education and earnings of their parents and about earnings of their friends if they knew it. The first year students were chosen, hence they are very close to the point of decision whether to start working with a secondary degree only or to postpone their earnings and go to the university. With most of them, one can expect they chose the tertiary education for additional gains (higher income) in the future.

As was proved in many previous research studies, the rates of return to higher education are high enough to "compete" with other forms of investment and to be able to cover some form of tuition fee

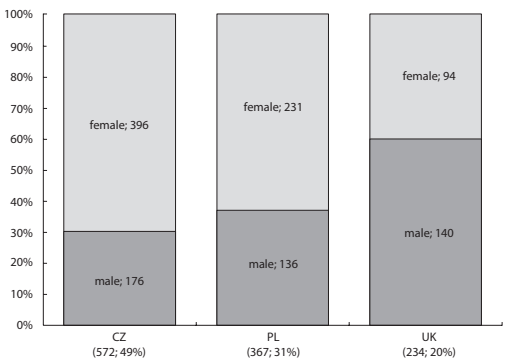
(see e.g. Belfield, 2000, Psacharopulous, 1995, or Psacharopulous, Patrinos, 2004). Discussions about implementation of tuition fees have been published in a number of papers and articles. An interesting approach is represented e.g. by Cahlík, T., et al. (2006) and can be taken for a good warning for proponents as well as opponents of tuition fee implementation.

The analysis described in this paper is based on the data from questionnaire surveys carried out during autumn 2009 at several faculties of economics at universities in the Czech Republic (Liberec, Pardubice and Prague), in Poland (Lublin) and in England (Huddersfield). Each of these countries has been using a different system of financing for tertiary education. All the data were collected personally which ensured high feedback from the students. The data analysis in the paper is a part of a long-term survey that has started in 2001.

For the purpose of this article, there are four key factors ("agents") to classify the respondents by: country of the survey, gender of the respondent, respondent's awareness of his/her friends' income, and the highest level of education of respondent's parents.

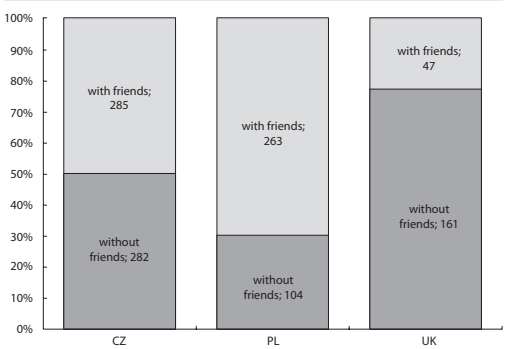
The distribution of respondents among the three countries is not uniform. Almost a half of all the respondents study at the Czech universities (CZ; 572 students; 49 per cent of all), less than one third were Polish (PL; 367 students; 31 per cent) and about one fifth were English respondents (UK; 234 students; 20 per cent). This imbalance is a consequence of a limited number of cooperating institutions outside the Czech Republic and will be taken into account during the research.

Figure 1 Classification of Respondents by Their Gender in 2009



Source: Authors' calculations, Survey of expectations 2009

Figure 2 Classifications of Respondents by Their Awareness of Their Friends' Income



Source: Authors' calculations, Survey of expectations 2009

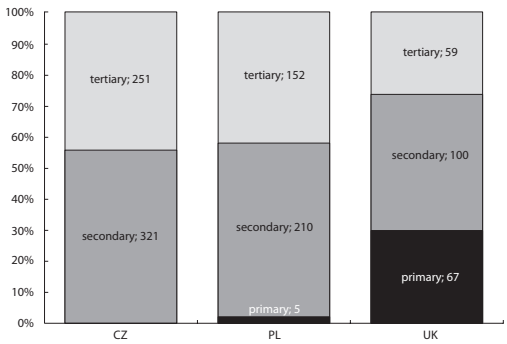
The major differences in the distribution of respondents by gender are shown in Figure 1. At the English faculty of economics the majority of students are men, while in the Czech Republic as well as in Poland, the young women constitute more than two thirds of students.

Rather surprising is the classification of respondents to those who are familiar with income of their friends and to those who are not. In the Czech Republic, the respondents are divided into these two groups almost half-and-half, while almost three quarters of Polish students have the information about salaries of their friends. On the other hand, nearly 80 per cent of English respondents did not admit they are familiar with incomes of their friends (see Figure 2).

The classification of respondents by the highest level of education of their parents seems also very interesting. The share of students coming from families where at least one of the parents earned university degree is in the Czech Republic and also in Poland higher than 40 per cent, while in the England, it is less than 30 per cent (see Figure 3).

One has to be very careful when making general judgments, since the number of English and also Polish respondents is much lower than the number of Czech ones and only one university

Figure 3 Classifications of Respondents by the Highest Level of Education of Their Parents



Source: Authors' calculations, Survey of expectations 2009

in the UK and one university in Poland took part in the survey. Therefore, generalizing about English or Polish students and their income expectations may not be as robust as for the Czech Republic. However, it is not the main task of this article to make conclusions about the English or Polish tertiary education students, but rather about the Czech ones. The English and Polish respondents will serve only for an etalon here.

The authors of this paper intend to calculate the rate of return on each year of the tertiary education (section 2) and then test the results of sensitivity to a couple of "agents" (such as gender, information about income of respondents' friends, and education of respondents' parents) that could possibly affect the rates of return the respondents expect to receive from their university degree. These tests will be performed in following sections. Their respective conclusions will be summed up in the last part of the paper.

2 RESEARCH METHODS

Following the method published by Psacharopoulos and Patrinos (1995, 2004), the expected returns on investment to tertiary education can be calculated using the formula of the short-cut method (1):

$$r = \frac{W_N - W_{wN}}{t \times W_{wN}}, \tag{1}$$

where r is the expected percentage rate of return on investment to tertiary education, W_N represents the expected income immediately after completing the university studies, W_{wN} stands for expected income the respondent would earn without the university degree, and t is a number of years of tertiary education ($t = 5$ for a master degree).

The main presumption for using such an equation is the constant shape of income curve for each respondent. It is almost certainly an overgeneralization and oversimplification, but authors of this article hope, for the purposes of this paper, this method is fairly justifiable and very useful especially for its clarity and easiness.

Nevertheless, some minor modifications have been done to calculate more precise and more easily interpretable rates of return. The equation (2) is using rather geometric than arithmetic average, as it is more suitable to determine the average annual rate of return of continuous valuation of expected incomes. Still, the main philosophy behind remains the same. For the purpose of the paper, we use the same equation even for England, abstracting from the fact, the English students have to pay tuition fees after finishing their studies. The range of tuition fee is dependent on the income of the former student. Therefore, the data available do not provide the information of the real value of tuition fees paid by each respondent. What will be calculated for the UK is not the real expected rate of return on investment in tertiary education, but a model expected rate under the circumstance of no tuition fees. On the other hand, it is also necessary to consider the question if the students' perception does or does not include the fact that they probably will have to pay some fee in the future (results including tuition fees in formula see Anchor et al., 2011, where level of tuition fees of English students expectations was included).

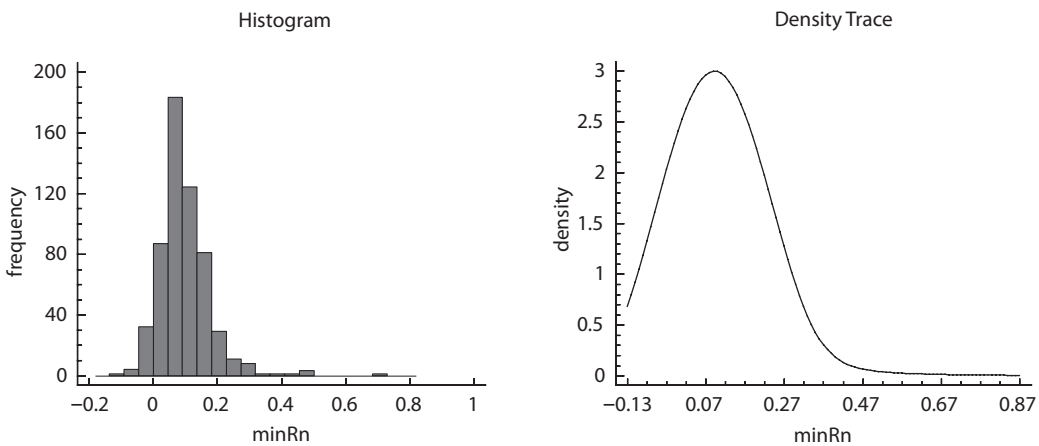
$$r = \sqrt[t]{\frac{W_N}{W_{wN}}} - 1. \tag{2}$$

Since the data received from the survey enable such a procedure, not only one rate of return on investment to tertiary education (*r*) was calculated for each respondent. The authors used minimal (the lowest), mean (the most probable), and maximal (the highest) values of expected spot incomes with and without university degree and calculated three levels expected spot rates of return: minimal (*minRn*), mean (*averRn*), and maximal (*maxRn*). Utilizing the expectations of respondent about their income in ten-year perspective (again with and without university degree), also the rates of return with ten-year-long working experience have been constructed – again at three levels: minimal (*minRt*), mean (*averRt*), and maximal (*maxRt*).

With these six levels calculated for each respondent, the next natural step would be to aggregate the numbers for all the respondents. The basic and obvious option would be an arithmetic average, however, it can be easily demonstrated the arithmetic average of the rates of return is not very suitable measure of central tendency. There are at least three important reasons for rejecting the method of arithmetic average: Arithmetic average – unlike modus or median – is an "artificial" statistical value; it does not necessarily have to represent any real number from the data set. The value of arithmetic average is predisposed to be biased by outliers (Seger, Hindls, 1995).

As the Figure 4 illustrates on the example of the lowest expected spot rates of return (*minRn*) of Czech students, the distribution of their responses is far from normal. We have chosen the Czech respondents and variable *minRn* for an example since the number of data is in this case by far the largest and the distribution of them should be therefore most probably the nearest to the normal distribution. Arithmetic average, median, and modus will therefore probably record significantly different values.

Figure 4 Histogram and Density Trace of Expected Minimal Spot Rates of Return on Investment to Tertiary Education (CZ minRn)



Source: Authors' calculations, Survey of expectations 2009

This simple "eye-ball test" can be supported by calculating the rate of skewness of the data sets. The standardized Fisher's skewness γ (3) should according to Wuensch (2005) fall into the confidence interval of <–2; +2> , if the distribution was normal or not significantly different from normal. The summary statistics for all three countries and all six calculated variables are shown in Table 1.

Table 1 Summary Statistics of the Data Sets

	minRn	*averRn*	*maxRn*	*minRt*	*averRt*	*maxRt*
Czech Republic						
Count	567	563	563	561	558	548
Average	0.099809	0.099242	0.106710	0.115353	0.119497	0.161108
Median	**0.084472**	**0.084472**	**0.095654**	**0.101972**	**0.107566**	**0.138633**
Mode	0.084418	0.107566	0.148698	0.148698	0.148698	0.148698
Skewness	21.3154	36.8485	28.8759	12.2525	16.1112	22.878
Poland						
Count	367	366	367	364	364	365
Average	0.107441	0.105439	0.149025	0.143985	0.145008	0.194043
Median	**0.084472**	**0.089613**	**0.110953**	**0.139723**	**0.118427**	**0.148698**
Mode	0.084472	0.107566	0.107566	0.148698	0.148698	0.148698
Skewness	14.7371	18.2828	22.3944	20.5429	23.0698	37.6907
United Kingdom						
Count	208	200	201	198	193	184
Average	0.105404	0.091221	0.067757	0.134843	0.134002	0.151379
Median	**0.089977**	**0.084472**	**0.069610**	**0.119364**	**0.118427**	**0.135471**
Mode	0.084472	0.84472	0.107566	0.148698	0.107566	0.148698
Skewness	8.15915	3.33054	–1.92432	9.1898	5.90539	6.0375

Source: Authors' calculations, Survey of expectations 2009

$$\gamma = \frac{\sum_{i=1}^{N} (X - \mu)^3}{\sigma^3 \times \sqrt{6 \times n}}, \tag{3}$$

where n is the count of the data set, X_i are the data values, μ is the average of these data, σ is their standard deviation.

There are three remarkable moments resulting from the Table 1:

Only in one case (UK *maxRn*) the calculated skewness fell below zero, suggesting the distribution of the variable is skewed to the left. But in this sole case the value of the skewness fits the confidence interval and the difference between the average and median is insignificant. Mode of UK *maxRn* seems to be considerably higher than average and median.

The fact the values of the rates of return with ten-year-long working experience are generally higher than the values of the spot rates of return (see Figure 4) only proves how naïve is the above postulated presumption of constant income function. Nevertheless, the authors still believe the method described above is sufficient and will serve their purposes effectively.

One should not be surprised by the fact, average and median values of *minRn* in the UK is higher than *averRn* and the values of *maxRn* are the lowest of these three. This merely suggests the British respondents expect the university degree can bring the highest increase to the lowest expected incomes. The higher levels of income these respondents allow for, the lower the expected rate of return on investment to tertiary education.

As a result of the section 2, authors decided to use median for a mean value estimate of the rates of return on investment to tertiary education. Its advantages over arithmetic average and mode have been sketched by Figure 4 and Table 1. Substantial skewness of distributions of basically all the levels of rates of return significantly deviates the values of average from the values of median and mode. Median shows better and more useful interpretability than mode regarding the aims of this article. Since the median value of *minRn* and *averRn* for the Czech students of faculties of economics reaches approximately 8.45 per cent (and *maxRn* is by one percentage point higher), we can conclude that at least half

of the students expect their income after university graduation will be by no less than 50 per cent higher than without the master degree.[3] For mode would be such an interpretation impossible or very awkward.

This simple conclusion of the research will be tested in the following sections for its sensitivity on several "agents" (country and gender of respondents, information about incomes of respondents' friends, highest degree of education of respondents' parents). All of these "agents" will be tested for having a significant influence on variance and especially on median of the analyzed factors (*minRn, averRn, maxRn, minRt, averRt, and maxRt*).

Since the skewness of the distribution of nearly all the factors has been proved, it is obviously not possible to use statistical methods based on the presumption of normal distribution of the data (such as *F*-test of variance homogeneity or *t*-test of mean value equality). For highly skewed distributions, Levene variance check (Levene, 1960) has been the most suitable (unlike Bartlett test or Cochrane test), esp. with the Brown-Forsythe modification (Brown, Forsythe, 1974). Its test statistic with distribution $F(\alpha, k-1, N-k)$ is defined as (4).

$$W = \frac{(N-k) \times \sum_{i=1}^{k} N_i \times (Z_{i.} - Z..)^2}{(k-1) \times \sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2}, \tag{4}$$

where W is the test statistics, k is the number of different groups the samples belong to, N is the total number of samples, N_i is the number of samples in the i-th group, $Z_{ij} = |Y_{ij} - \tilde{Y}_{i.}|$, $Z_{i.}$ are the group means of Z_{ij}, $Z..$ is the total mean of all Z_{ij}, Y_{ij} is the value of the j-th sample from the i-th group, $\tilde{Y}_{i.}$ is the median of the i-th group, and α is the level of significance (here 5 per cent).

The variance check will be used to indicate whether the variance of the analyzed factors differs significantly when allowing for various "agents". But this test and its results are not of the central focus of the authors. The main task is to test the median sensitivity to different "agents" (median robustness). For this purpose, authors decided to use Mann-Whitney-Wilcoxon (MWW) median test as it is more sophisticated and robust than today rather obsolete Mood median test (Mann, Whitney, 1947). The MWW test statistic has for large samples approximately normal distribution. The formula of the test statistic can be written as (5):

$$U = n_1 \times n_2 + \frac{n_2 \times (n_2 + 1)}{2} - \sum_{i=n_1+1}^{n_2} R_i, \tag{5}$$

where n_1 and n_2 are the size of the samples and Ri are the ranks.

The standardized z value can be formulated as (6):

$$z = \frac{U - \mu_U}{\sigma_U}, \tag{6}$$

where the average of U is defined as:

$$\mu_U = \frac{n_1 \times n_2}{2}, \tag{7}$$

[3] If each year at the university yields 8.4472 per cent to the expected income, then after five years of master studies the students expect $(1 + 0.084472)^5 = 1.5$ higher income.

and the standard deviation of U can be written as:

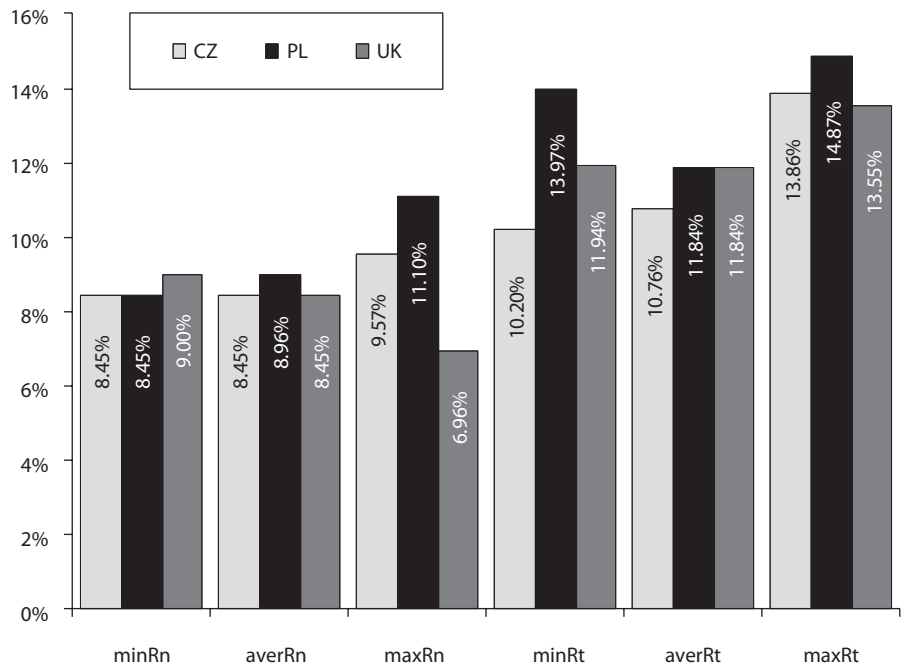
$$\sigma_U = \sqrt{\frac{n_1 \times n_2 (n_1 + n_2 + 1)}{12}}. \quad (8)$$

3 COUNTRY COMPARISON OF THE EXPECTED RETURNS ON INVESTMENT IN TERTIARY EDUCATION

The first question, the authors had asked after calculating the median values, was how significantly differ these expected rates of return on investment in tertiary education across the triplet of analyzed countries. Or: Is the country where the survey has been carried out an important "agent" affecting significantly the median value of the rate of return?

The country differences have been illustrated in the graph (see Figure 5), backed up by the Lavene and MWW statistical procedures.

Figure 5 Country Comparison of the Expected Rates of Return on Investment in Tertiary Education



Source: Authors' calculations, Survey of expectations 2009

The variances of all factors were found significantly lower in the Czech Republic against Poland, while variances of all Polish results proved to be significantly higher than in the United Kingdom. The UK *minRt* variance proved to be higher than the Czech one, while the UK *maxRt* variance was found lower than its Czech counterpart.

A quite interesting fact resulted from the median testing: The medians of *minRn*, *averRn*, and *maxRt* were not found significantly different. All three countries recorded similar values of medians of these three levels of expected rates of return. When testing the medians of *maxRn*, the British value was proved to be the lowest, the Czech one significantly higher, and the Polish even higher. The country also matters when discussing the values of *minRt* and *averRt*. The Polish respondents expect significantly higher *minRt*, while the British students are more modest in their expectations, and the Czech ones lag even

behind their British colleagues. And at last: The British and Polish respondents expect approximately the same rate of return on the level *averRt*, while the Czech students expect significantly less.

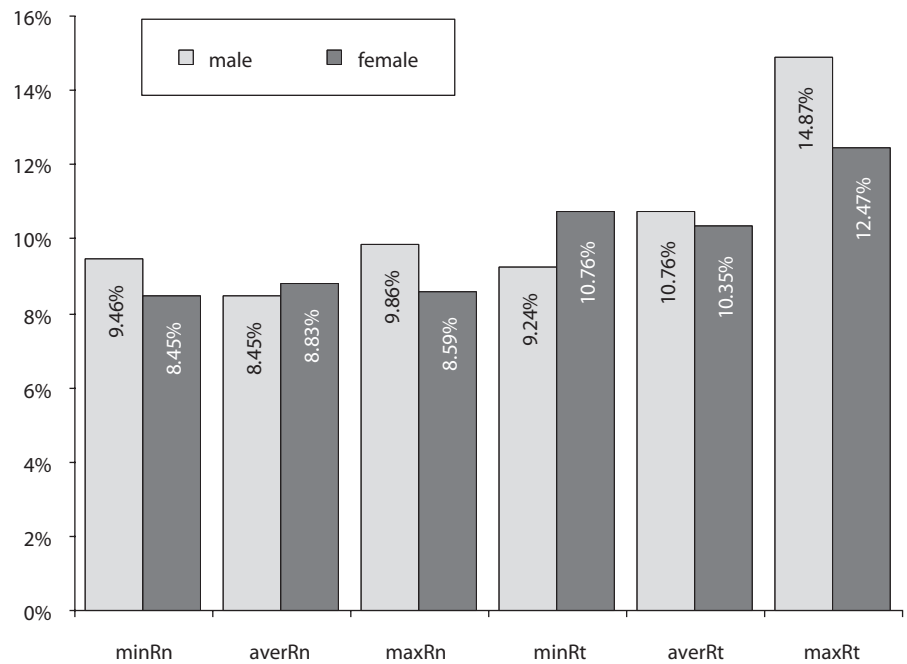
Obviously, one has to be very careful when generalizing this conclusion, especially because of the limited number of respondents in Poland and mainly in the UK, who – to make things even less satisfying – all come from one university. Yet, what seems undoubted is the fact, that minimal and mean expected rates of spot return do not differ significantly among the three countries (just the same can be concluded about the maximal expected rates of future return with ten-year-long working experience).

Authors took into account the restrictions and problems arising from smaller number of respondents and cooperating institutions in Poland and in the UK and decided to continue in their tests only with the data from the Czech Republic.[4]

4 SENSITIVITY OF EXPECTED RETURNS ON INVESTMENT IN TERTIARY EDUCATION TO THE GENDER OF RESPONDENT

Next question the authors had asked was focused on gender differences in the expected rates of return. It should be pointed out again the interest of the authors was not in the expected absolute values of in-

Figure 6 Sensitivity of Expected Rates of Return on Investment in Tertiary Education to the Gender of Respondent



Source: Authors' calculations, Survey of expectations 2009

[4] Authors carried out all the steps following in sections 4, 5 and 6 also with the data from Polish and British surveys, but the results (esp. the range of the confidence intervals) were negatively affected by the insufficient and incomparable numbers of respondents in particular subgroups which led to the conclusion to describe in this article mainly the results of the remaining sensitivity tests for the Czech Republic.

come, but in the expected relative increases due to university studies. The fact women expect significantly lower incomes even with the university degree was identified and tested in Urbánek et al. (2009).

An interesting thing to comment on may be the variance of expected rates of return in the Czech Republic. The Levene variance checks proved basically for all the analyzed factors significantly higher variance for male respondents than for female (with the only exception in *minRt*, in this case the variances do not seem to differ significantly).

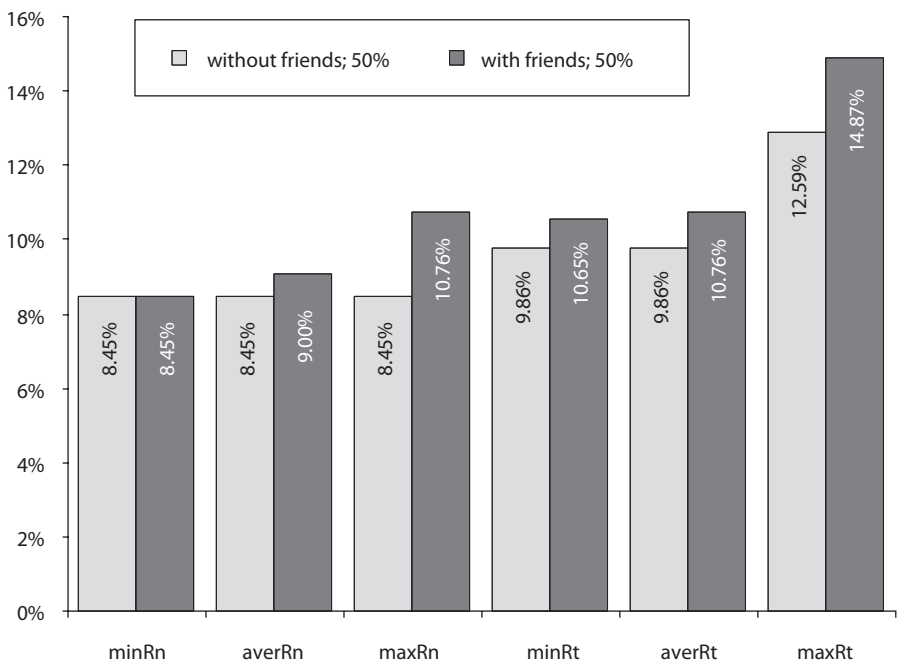
Nevertheless, the sensitivity of expected rates of return on the tertiary education to the gender of respondents was found surprisingly weak. The gender of Czech respondents does not make much difference and the diversity between the genders at the particular levels of rates of return (namely for *minRn*, *averRn*, *maxRn*, *minRt*, and *averRt*) was not found significantly strong. The only exception to this conclusion represents *maxRt*. In this case, the male respondents showed significantly higher imagination about their future incomes while females stuck closer to the ground (see Figure 6).

5 SENSITIVITY OF EXPECTED RETURNS ON INVESTMENT IN TERTIARY EDUCATION TO THE INFORMATION FROM RESPONDENT'S FRIENDS

The authors also suspected the information about friends' income situation may affect the expected rates of return (for broader analysis, see e.g. Urbánek et al., 2010). Analogically to the routine followed in section 4, authors divided the data set into two groups: one made up from respondents who submitted the information about their friends' incomes and the other one containing the rest of the respondents. Then, the authors ran again the testing procedures described in section 1.

The results of variance check are again the first interesting point worth mentioning. The information about the income of respondents' friends raises significantly the variance of the expected rates

Figure 7 Sensitivity of Expected Rates of Return on Investment in Tertiary Education to the Information from Respondents' Friends



Source: Authors' calculations, Survey of expectations 2009

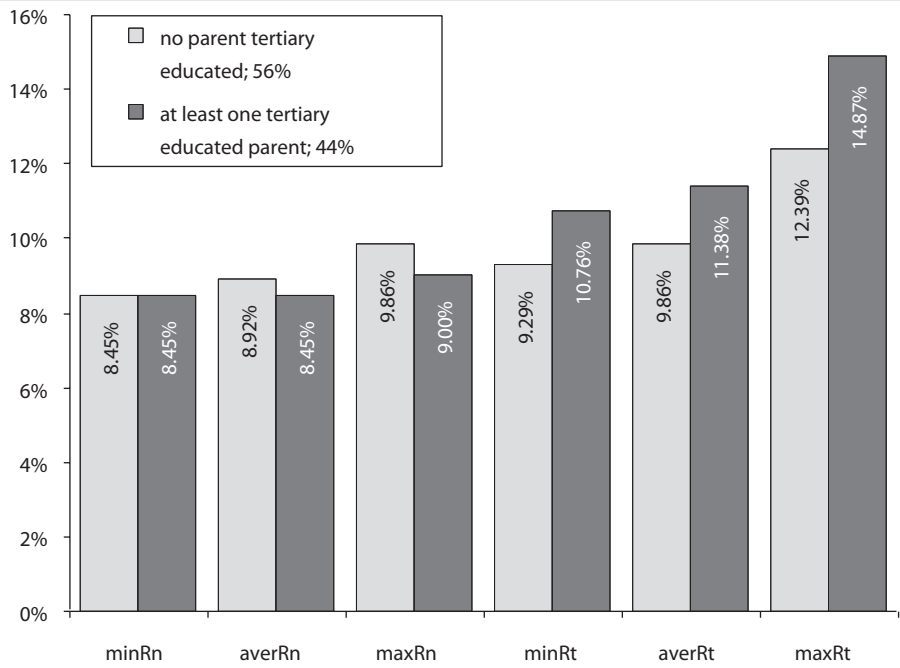
of return at basically all levels. The only exception to this is represented by averRn. In this case, the variances between the two groups of respondents (with and without knowledge on their friends' income) do not differ significantly.

Unlike the gender, the information about income of their friends seems to be a strong factor influencing the expectations of the respondents. While the lower levels of expected rates of return (*minRn* and *minRt*) were not proved to be significantly sensitive to the friends' income of respondents, the conclusions are quite the opposite for the higher levels (*averRn*, *maxRn*, *averRt*, and *maxRt*). The MWW tests discovered that respondents who are familiar with their friends' incomes tend to expect significantly higher returns on investment to tertiary education on the mean and maximal levels not only on spot but also after ten years of working experience. The minimal expectations are not affected by the information from friends to evincible extent (see also Figure 7).

6 SENSITIVITY OF EXPECTED RETURNS ON INVESTMENT IN TERTIARY EDUCATION TO THE DEGREE OF EDUCATION OF THE RESPONDENTS' PARENTS

The last testing section of the paper discusses the sensitivity of expected rates of return to the highest degree of education achieved by any of the parents. The respondents were again divided into two samples, one containing those whose parents (at least one of the parents) received the university degree, the other sample consists of the rest of the respondents (none of their parents has university degree). The differences between the two samples measured by the median of the expected rate of return are demonstrated by Figure 8.

Figure 8 Sensitivity of Expected Return on Investment in Tertiary Education on the Highest Level of Education of the Respondents' Parents



Source: Authors' calculations, Survey of expectations 2009

The results of the MWW tests proved once again the robustness of the conclusions in the section 2 only partially. The spot expected rates of return (*minRn*, *averRn*, *maxRn*) remained unbiased by the level

of education of the parents, while the expected rates of return after ten-year-long working experience were recorded significantly higher for respondents with tertiary educated parents. The Levene variance checks on the other hand discovered, the responses of students with at least one tertiary educated parent tended to record significantly higher variance at the mean and maximal levels of expected rates of return (spot as well as with ten-year-long working experience, i.e. *averRn*, *maxRn*, *averRt*, and *maxRt*).

CONCLUSION

The findings of this paper and other studies (see e.g. Anchor, 2011, Psacharopoulos, 1995 etc.) indicate that there is a significant expected pay off to higher education. Moreover, the expected returns increase with work experience which suggests that the benefits from higher education are larger in the medium term than immediately after graduation. To discuss financial participation of individuals at public tertiary education, it is crucial to know the earnings expectations of university students. All the tests performed in the paper showed and proved one critical piece of information: the values of the spot expected rates of return on investment to tertiary education calculated for the Czech Republic never fell below 8.45 per cent (and the rates with ten-year-long experience never fell below 9.29 per cent). Whatever the gender of respondents, whatever information about their friends' incomes they dispose of, whatever the degree of their parents' education, the majority of Czech students at faculties of economics expect their income will increase at least by 50 per cent[5] after they receive their master degree. This conclusion supports also the findings of Filer et al. (1999) who calculated very similar values of the expected rates of return on investment to tertiary education twelve years ago, although they used a rather different methodology. The Czech results also do not differ much from the outcomes of surveys in Poland and in the United Kingdom.

The results show that students expect a higher wage premium to compensate for the perceived costs. We can argue if even tuition fees (also deferred fees), which will eventually act as a disincentive to enter higher education since students will not expect indefinitely that their future employers will be able to offer them a wage premium high enough to compensate for the expected costs of higher education (for results see Anchor et al., 2011, Barr, 2010). Private participation on financing of tertiary education should not be meant as a response to fiscal constraints only. It is necessary to consider a parallel microeconomic argument: the tertiary education has significant private benefits, justifying a contribution from the beneficiary on both efficiency and moral grounds. Thus the case for some private finance might be robust, but policy needs to be designed carefully so that it does not harm efforts to widen participation and does not discriminate people from poorer social conditions.

ACKNOWLEDGEMENTS

The data used in the paper were collected with the support of the Czech Science Foundation, No. 402/09/1123.

References

ANCHOR, J., FIŠEROVÁ, J., MARŠÍKOVÁ, K., URBÁNEK, V. Student Expectations of the Financial Returns to Higher Education in the Czech Republic and England: Evidence from Business Schools. *Economics of Education Review*, 2011, 4 (August), pp. 673–681.

BARR, N. Alternative Funding Resources for Higher Education. *Economic Journal*, 1993, Vol. 103, pp. 718–728.

BARR, N. *Paying for Higher Education: What Policies, in What Order?* [online]. London School of Economics, Submission to the Independent Review of Higher Education Funding and Student Finance, 2010, pp. 9–10. [cit. 21.3.2011]. <http://econ.lse.ac.uk/staff/nb/BarrJohston_Interestsubsidies100215.pdf>.

[5] This result is based on the most moderate expectations of majority (or at least one half) of Czech respondents, i.e. on the annual rate of return at 8.45 per cent for every year of master studies.

BECKER, G. S. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* Chicago: University of Chicago Press, 1993.

BELFIELD, C. R. *Economic Principles for Education, Theory, and Evidence.* Northampton, MA, USA: Edward Elgar Publishing, Ltd., 2000.

BROWN, M. B., FORSYTHE, A. B. Robust Tests for the Equality of Variances. *Journal of the American Statistical Association,* 1974, Iss. 69, pp. 364–367.

CAHLÍK, T., HLAVÁČEK, J., MARKOVÁ, J. Školné či dotace? Simulace a modely systému vysokých škol (Tuition or Grants? Simulations and Models of Higher Education Systems). *Politická ekonomie*, 2006, Iss. 1, pp. 54–66.

FILER, R. K., JURAJDA, Š., PLÁNOVSKÝ, J. Education and Wages in the Czech and Slovak Republic during Transition. *Labour Economics*, 1999, Vol. 6, Iss. 4, pp. 581–593. ISSN 0927-5371.

LEVENE, H. Robust Tests for Equality of Variances. In OLKIN, I. et al. (Eds.) *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling.* Palo Alto, CA, USA: Stanford University Press, 1960, pp. 278–292.

MANN, H. B., WHITNEY, D. R. On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 1974, 1, pp. 50–60.

MATĚJŮ, P. et al. Bílá kniha terciárního vzdělávání (White Book on Tertiary Education)[online]. Prague: Ministry of Education, Youth and Sports, 2009. [cit. 8.5.2009]. <http://www.msmt.cz/uploads/bila_kniha/schvalena_bktv/BKTV_finalni_verze.pdf>.

PSACHAROPULOUS, G., PATRINOS, H. A. Returns to Investment in Education: A Further Update. *Education Economics*, 2004, Iss. 2, pp. 111–134.

PSACHAROPULOUS, G. The Profitability of Investment in Education: Concepts and Methods [online]. *Human Capital Development and Operations Policy Working Papers.* 1995. [cit. 30.6.2010]. <http://commdev.org/files/1863_file_Profitability_Investment_Edu.pdf>.

ŘEHOŘOVÁ, P., MARŠÍKOVÁ, K., URBÁNEK, V. Policy in Higher Education Development: Return on University Investment and Tuition Fees. E+M *Ekonomie a Management,* 2010, Vol. 8, Iss. 1, pp. 37–45.

SEGER, J., HINDLS, R. *Statistické metody v tržním hospodářství* (Statistical Methods in Market Economy). Prague: Victoria Publishing, 1995.

ŠÁDEK, M. *Analysis of Investment in Higher Education in Selected European Countries.* Diploma Theses. Liberec: Technical University of Liberec, 2010.

URBÁNEK, V., MARŠÍKOVÁ, K., ŘEHOŘOVÁ, P. *Návratnost investice do vysokoškolského vzdělání: komparace očekávaných a reálných výdělků I* (The Return on Investment in the Higher Education: A Comparison of Expected and Actual Earnings I). Liberec: Technical University of Liberec, 2009.

WOLTER, S. C., WEBER, B. A. A New Look at Private Rates of Return to Education in Switzerland. *Education + Training*, 1999, Iss. 8, pp. 366–372.

WUENSCH, K. L. Skewness. In EVERITT, B. S., HOWELL, D. C. (Eds.) *Encyclopedia of Statistics in Behavioral Science*. New York, USA: John Wiley & Sons, 2005, pp. 1855–1856.

Multivariate Statistical Analysis of the E-Communication Indicators in the European Union

Julie Poláčková[1] | *Czech University of Life Sciences in Prague, Czech Republic*
Zuzana Pacáková[2] | *Czech University of Life Sciences in Prague, Czech Republic*

Abstract

The aim of the paper is to analyze the e-communication in the member states of the European Union. On the basis of data from the Eurobarometer 75.1 survey and from the Eurostat database the differences in e-communication level among the European Union states were analyzed. Principal component analysis was used for the general analysis of differences between the states of European Union. It was possible to explain 77% of the total variance by the first two components. The first component represents the level of e-communication while the second component characterizes quality of services and proportion of advanced Internet users. The overall e-communication level was evaluated using the component indicator. The Northern states together with the Netherlands and Luxembourg achieve the highest e communication level. On the other hand, the southern states (Greece, Italy, Portugal) together with the two new member states (Romania, Bulgaria) received the lowest rating, while Romania has been found an outlier on the basis of principal component analysis as well.

Keywords

E-Communication, households with internet, internet and computer skills, principal component analysis, composite indicators

JEL code

C38

INTRODUCTION

Computer and Internet education is a key prerequisite of the information access. Knowledge of information and communication technologies, also called „computer literacy", constitutes basic requirement for further development of individuals in modern world. This claim is based on the assumption that about a half of productivity growth in modern economies is related to the use of information and communication technologies (see EDCL, 2012).

[1] Department of Statistics, Faculty of Economics and Management, Czech University of Life Sciences in Prague, Kamýcká 129, 165 21 Prague 6 – Suchdol, Czech Republic. Corresponding author: e-mail: polackova@pef.czu.cz, phone: (+420)224382299.
[2] Department of Statistics, Faculty of Economics and Management, Czech University of Life Sciences in Prague, Kamýcká 129, 165 21 Prague 6 – Suchdol, Czech Republic.

Technological advancement has totally changed the field of communication. Information and communication technology (ICT) has become ever more ubiquitous throughout society. The Internet is now the center of economic, cultural and political life. It is used as a mechanism for delivery of public services, personal communication, and as a vast source of information and entertainment. A detailed analysis of this issue is essential for measuring and comparing computer literacy across the EU. The results can be used to set a policy of information education.

The main aim of this paper is to analyze the level of e-communication in different countries of the European Union. The analysis combines information obtained from the Eurobarometer survey and from the accessible resources of Eurostat. The major issue was to analyze the frequency of access to the Internet and its usage by individuals. In addition, the quality of Internet connection in households was examined. This is an important factor for evaluating the effectiveness of access to information. Another examined indicator was the usage of mobile connectivity as a modern e-communication tool.

Detailed analysis was based on the important indicators related to e-communication. These refer to indicators related to ICT equipment (proportion of individuals having computer, proportion of households with computer connected to the Internet, proportion of individuals having mobile Internet, or proportion of broadband penetration rate), quality of services (mobile phone satisfaction index and internet satisfaction index), and users´ knowledge and skills (computer and Internet skills, proportion of individuals using internet banking or e government, proportion of individuals making phone calls over the Internet, or ordering goods over the Internet).

1 INFORMATION AND COMMUNICATION TECHNOLOGIES

Technological advancement is a permanent feature of human society. However, this advancement is by its nature an inconsistent phenomenon, with periods of relative stability in technological capability punctured by periods of rapid innovation that can have profound consequences for society. Berry (2011) points out that in the field of information technology have been rapid developments in computing since the post-war period. Computers have made an immense progress in terms of processing power, speed and capacity in recent decades. In communication, the progression aimed at ever more sophisticated and flexible means of communication, from the telegraph to the mobile telephone. Over the last 30 years these two spheres have largely ceased to develop separately. But it is the synergy between information and communication technology that has been most revolutionary.

The most remarkable way of the biggest technological advancement is the World Wide Web. This technology began its life as a relatively simple means to enable communication between linked computers, and has evolved into ever more diverse forms, from websites, e-mail and social networking to telephony, video streaming and interactive virtual worlds. This technology is used in countless varieties of form. Among other uses, it is a mechanism for the delivery of public services, a means of personal communication, a vast source of information, knowledge, and entertainment, and a tool allowing for new industrial practices (Berry, 2011).

The computer literacy is a part of information literacy. Recent studies (see Dombrovská, 2011, Dohnálková, Landová, 2009) in this context refer to the information literacy, which consists of document, literary, linguistic, numeric and computer literacy. American Library Association (1989) defines information literacy as a set of abilities requiring individuals to "recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information". Information literacy is also increasingly important in the contemporary environment of rapid technological change and proliferating information resources. It is related to information technology skills, which enable an individual to use computers, software applications, databases, and other technologies to achieve a wide variety of academic, work-related, and personal goals. Information literacy, while showing significant overlap with

information technology skills, is a distinct and broader area of competence. Increasingly, information technology skills are interwoven with, and support information literacy.

Recent studies show that main determinants of the intensity of Internet use in Europe are education and household income (for details see Montagnier, Wirthmann 2011). Differences between the countries can be caused by the different approaches of governments to the support and expansion of ICT. Issue of accessing the Internet by older people and excluded social groups can be found for example in Berry, 2011.

2 DATA SET AND METHODS

E-commmunication of the EU states was characterized by a set of fourteen indicators (see Table A1 in the Annex). Each indicator represents the percentage of population with a specified characteristic (having computer, using Internet banking etc.). The data set used in the e communication analysis was obtained from the Eurobarometer 75.1 survey (for details see European Commission, 2011) and from the Eurostat data source: section Information society. The statistics in this section track the usage of information and communication technologies (ICT). The methodological manual for surveys on ICT usage can be found on Eurostat web pages (Eurostat, 2012). The aggregated variables taken from Eurostat database refer mainly to the year 2011, only the *Broadband penetration rate, E-government usage and Internet banking usage* were available just for the year 2010. The Eurobarometer 75.1 was realized also in 2011 (February – March). It is particularly focused on e-communication in the household: mobile phone, television and Internet. In all, Eurobarometer 75.1 interviewed 26 836 citizens in the 27 countries of the European Union. All respondents were residents in the respective country, nationals and non-nationals but EU-citizen aged 15 and over. A multi-stage, random (probability) sampling design was used for this Eurobarometer. For the purpose of the further analysis the *WEIGHT EU27 (W22)* was used. It adjusts each sample in proportion to its share in the total population of the European Union aged 15 and over. It includes all 25 member countries after the 2004 enlargement, and the new members as of 2007 Romania and Bulgaria (for details see European Commission, 2011).

The Eurobarometer data set contains computed satisfaction indexes. These indexes are presented as discrete variables at four point ordinal scale. The mobile Internet satisfaction index was computed from the following questions: *mobile phone never cuts-off, it is always able to connect, user doesn't limit calls due to charges, and user doesn't limit mobile Internet due to charges.* The Internet satisfaction index was based on questions: *connection never breaks down, speed matches contract conditions, and the provider's support is useful.*

The indicators are presented on a six point ordinal scale in the Eurobarometer survey. For the purpose of further analysis the responses of individual respondents were aggregated. The proportion of positive responses in each state was used in following computations. Also the proportions of positive responses of aggregated indicators from the Eurostat database were used.

2.1 Principal component analysis

By reducing a data set from a group of related variables into a smaller set of components, the principal component analysis (PCA) achieves parsimony by explaining the maximum amount of common variance using the smallest number of explanatory concepts.

The original variables x_i, $i = 1, ..., m$, can be reduced to a smaller number of principal components y_j. The principal components are uncorrelated linear combinations of the original variables. All linear combinations are related to other variables or to the data structure. The principal components explaining the maximum amount of variance of the original variables (for details see Hebák et al., 2007, Rencher, 2002, or SAS Documentation, 2008). The first principal component corresponds to the direction of maximum variance; the second principal component corresponds to the direction of maximizing the

remaining variance, and so on. Each principal component corresponds to a certain amount of variance of the whole dataset.

For the purpose of constructing the composites indicators the unstandardized principal component scores were normalized according to the equation:

$$\gamma_r = \omega_r \sqrt{\lambda_{(r)}}, \tag{1}$$

where γ_r is the normalized vector of component loads, ω_r are eigenvectors and $\lambda_{(r)}$ represent the eigenvalues (Hebák et al., 2007).

The further useful dimension reduction device is to evaluate the first two principal components for each observation vector and to construct a biplot. Biplot allows information on both samples and variables of a data matrix to be displayed graphically. Samples are displayed as points while variables are displayed either as vectors. A biplot is an enhanced scatterplot that uses both points and vectors to represent a structure. As used in principal component analysis, the axes of a biplot are a pair of principal components. A biplot uses points to represent the scores of the observations on the principal components, and it uses vectors to represent the coefficients of the variables on the principal components. For details of this application, see Meloun et al. (2006), or Rencher (2002).

For the purpose of this analysis the SAS 9.2 software was used to construct the principal components and related plots. The standardized principal components were computed in MS Excel according to the Equation 1.

The PRINCOMP Procedure was used to fit a principal component model. Also the PRINQUAL Procedure was used to fit a model with optimal linear transformation of the variables and for graphical results. The advantage of the PRINQUAL procedure is that results contain a biplot.

2.2 Composite indicators

The final part of the paper is focused on constructing composite indicators (CI). The composite indicators are widely used as a tool providing a simple but complex comparison of countries and regions. The most important advantages of composite indicators are the following: easier interpretation of the results compared to a set of indicators and also a dimension reduction without a loss of information (for detailed discussion see OECD, 2008).

To accomplish the different variables comparability the sub-indicators which are summarized by the use of CI have to be normalised first. There are several methods ensuring the data comparability. The selected method of normalisation should take into account the properties of the data – respect to the measurement units in which the indicators are expressed and the robustness against possible outliers in the data (see Ebert, Welsch, 2004).

The original values were expressed as a ratio to the median value. In case of such indicators where the lower value indicates better position of the state, the ratio was expressed inversely. The normalised values of all sub-indicators were then aggregated using the arithmetic mean.

The unweighted composite indicator for the i-th state is computed using the following formula:

$$CI_i = \frac{\sum_{j=1}^{p} y_{ij}}{p}, \tag{2}$$

where p is the number of indicators, y_{ij} is the ratio of the original value to the median value of the *j-th* indicator computed as $y_{ij} = x_{ij} / \tilde{x}_{.j}$, where x_{ij} is the original value for the *i-th* state and the *j-th* indicator, i=1, 2, …, 27 and j=1, 2, …, p and $\tilde{x}_{.j}$ is the median value of the *j-th* indicator.

The weighted composite indicator was computed using following formula:

$$wCI_i = \sum_{i=1}^{p} y_{ij} w_j, \tag{3}$$

where y_{ij} is the ratio of the original value to the median value of the *j-th* indicator and w_j is the weight of *j-th* indicator, j=1,2, …., p and $\sum_{j=1}^{p} w_j = 1$, p is the number of indicators. There are several weighting methods used for the construction of a composite indicator (see OECD, 2008, Saisana, 2002, Munda, Nardo, 2005). In case of correlation among the sub-indicators Saisana (2002) recommends to use the principal components with the objective of combining sub-indicators into composite indicators to reflect the maximum possible proportion of the total variation in the data set.

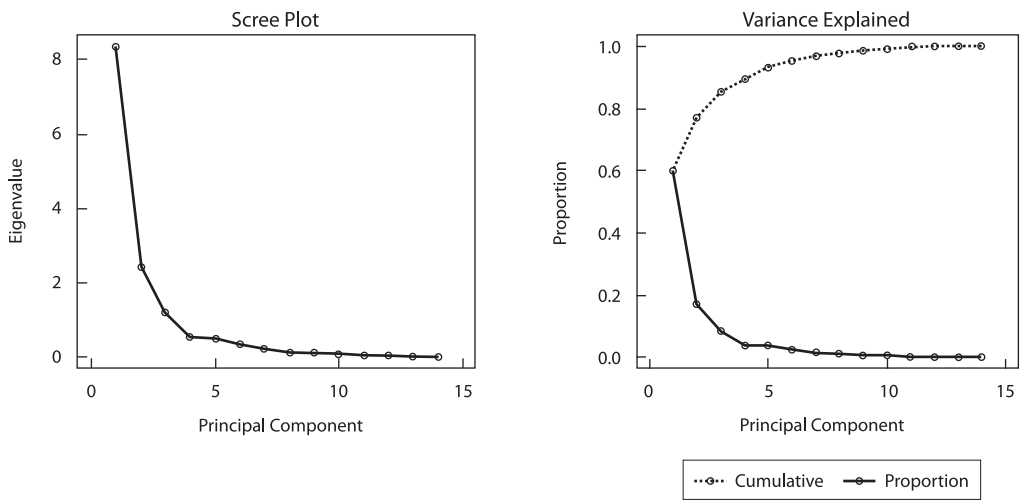
The composite indicators were computed in MS Excel as well as the construction of related figures.

3 DATA ANALYSIS AND RESULTS

The principal component analysis (PCA) was used for the general analysis of differences between the states of European Union. PCA identifies patterns in data and highlights their similarities and differences according to a varied level of e-communication.

In PCA, we seek to maximize the variance of a linear combination of the input variables. The first two principal components account for 77% of variance of the whole dataset and the first three account for 85%. Figure 1 shows a scree plot and a plot of cumulative proportional eigenvalues. The eigenvalues indicate that two or three components provide a good summary of the data.

Figure 1 Scree plot of eigenvalues and proportion of explained variance plot



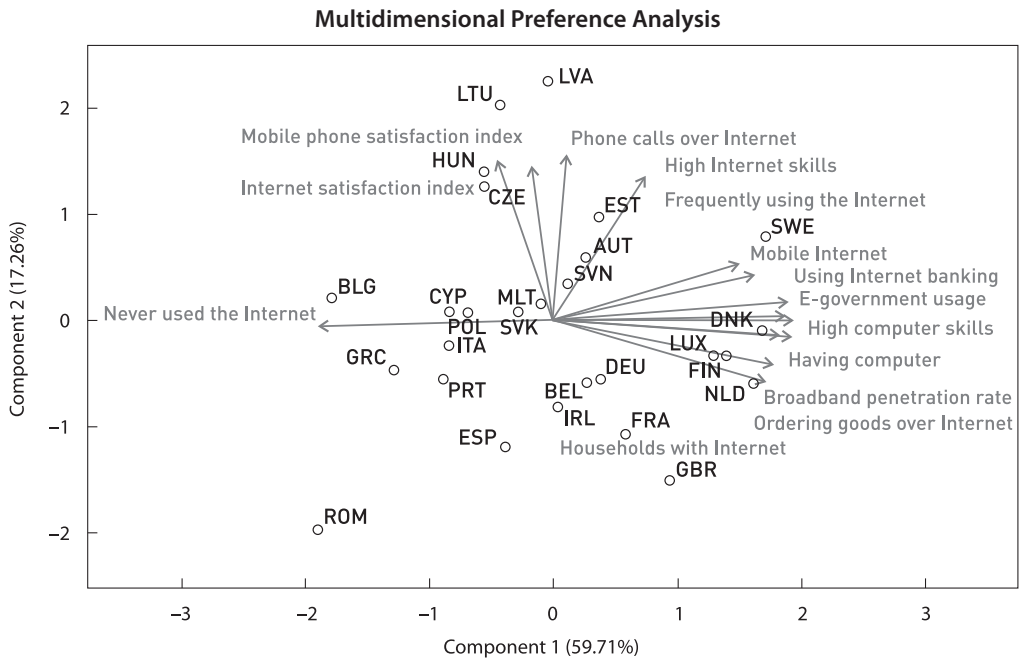
Source: Own construction

The first principal component is the linear combination with maximal variance. It explains about 60% of the total dataset. It largely represents 10 input variables, which are logically related. The corresponding eigenvector expresses an association of input variables with the first principal component (see Table 1). The second principal component accounts for 17% of variance and it has high positive loadings on four indicators.

The eigenvalue of the third component is 1.18 and it accounts for 8% of the total variance. The interpretation of the third component is not obvious, therefore it will not enter into following computations. Subsequent components contribute less than 5% of the total variance each. For the purpose of further analysis two dimensions are sufficient.

The objective of the following analysis is to evaluate the first two principal components for each observation vector and to construct a biplot. Biplot uses points to represent the scores of the observations on the principal components, and it uses vectors to represent the coefficients of the variables on the principal components. A vector points in the direction, which is the most similar to the variable represented by the vector. This is the direction which has the highest squared multiple correlation with the principal components. The length of the vector is proportional to the squared multiple correlation between the fitted values for the variable and the variable itself. Vectors that point in the same direction correspond to variables that have similar response profiles, and can be interpreted as having similar meaning in the context set by the data. Figure 2 shows the constructed biplot of the first two principal components. Vectors represent questions, and the points represent states. A group of vectors pointing in the same direction correspond to a group of questions, which have a similar proportion of positive answers across all states.

Figure 2 Biplot of the first two principal components



Source: Own construction

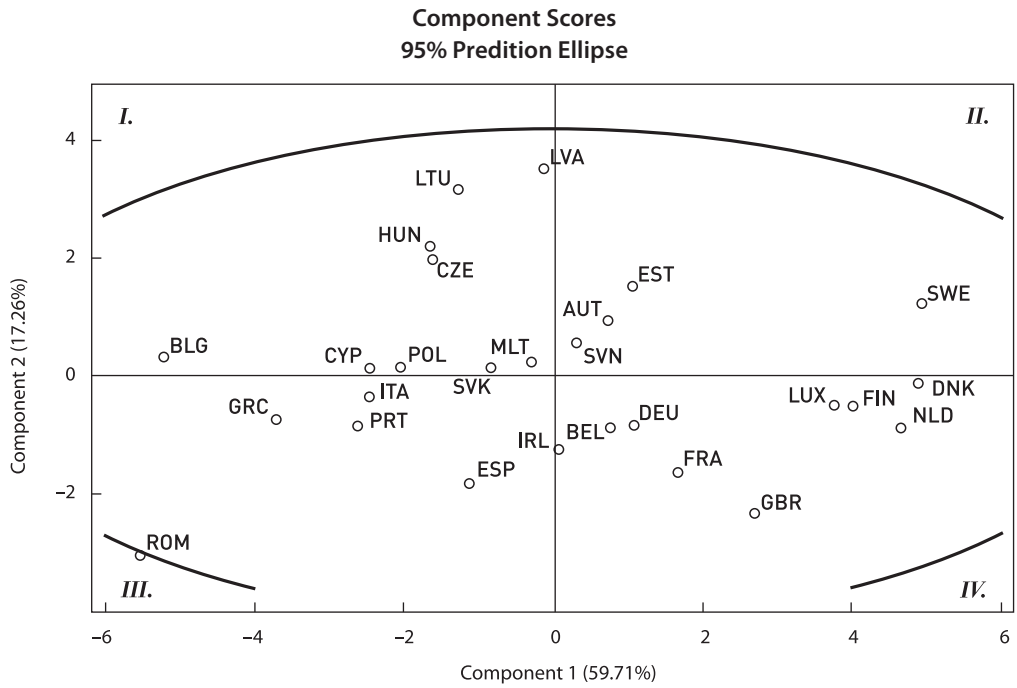
The first principal component has high negative loadings on variables *Never used the Internet* and high positive loadings on 9 input variables related to equipment and Internet use. Therefore it is obvious that the higher component score of this component means a higher level of e-communication in the country.

The second principal component represents four input variables. This component is correlated with indicators of the quality of services (mobile phone and Internet satisfaction index), and also with variables *Phone calls over Internet and High Internet skills*. It refers to the relationship between the level of the quality of services and the proportion of advanced Internet users.

Figure 3 displays the component score plot with a 95% prediction ellipse overlaid. It shows the spread of individual observations in the first two dimensions. The points that are close together correspond to observations that have similar scores. It is possible to identify regional trends on this plot or identify Romania as a possible outlier.

The plot can be split into 4 different quadrants. First quadrant represents a low level of household's equipment and Internet use and a high level of the quality of services. It includes mostly the countries of Central and Eastern Europe, such as Czech Republic, Hungary, Lithuania, Latvia, Bulgaria, Cyprus, Poland, Slovakia, and Malta. Although the quality of services provided by individual operators and other companies in these countries are usually not achieving the level of western countries, the users are mostly satisfied with it.

Figure 3 Component score plot with a 95% prediction ellipse



Source: Own construction

The second quadrant, which represents the highest level of e-communication (the highest level of ICT usage, and the quality of services) includes four states only: Slovenia, Austria, Estonia and Sweden. The first three of these states are located closer to the center of the plot, while Sweden reached the highest value of the component score in the 1st components. Together with Denmark, the Netherlands, Finland and Luxembourg it has the highest level of e-communication usage and equipment.

The third quadrant represents a low level of e-communication. It contains coastal states: Greece, Italy, Portugal, Spain, and Romania with the lowest level at all. With the exception of Romania, there are western countries with developed agriculture and tourism – people working on farms or in accommodation services do not see the Internet and electronic communication as a significant contribution to their work.

The fourth quadrant includes states with high level of ICT usage, but also with high requirements to the quality of services. Residents of the states with those services are rather dissatisfied, which may not mean lower standards than in other EU countries, but rather the higher demands of users. This quadrant represents the most advanced EU member states, such as: Ireland, Belgium, France, Great Britain, Luxembourg, Finland, Denmark, and Netherlands.

3.1 Composite indicator

To evaluate the overall e-communication level among the EU states a composite indicator (CI) was constructed. The original values of all fourteen sub-indicators were expressed as a ratio to the median value. While there are wide differences among the EU states from the view of selected ICT indicators the median was used instead of the mean which is sensitive to extreme values. The ratio then reflects the position of each state while higher values indicate better position of the state (the values above one indicate the position above the median value). In case of such indicators (e.g. *percentage of people who never used Internet*) where the lower value indicates better position of the state, the ratio was expressed inversely.

The normalised values of all sub-indicators were then aggregated using the arithmetic mean. So the values close to one indicate that the state is in average close to the median value of the EU states on the basis of the selected indicators. The higher is the value of the composite indicator, the better is the position of the state.

The composite indicator of the e-communication level was constructed in both unweighted and weighted forms, while the weights were based on the results of the principal component analysis.

The weight was set down on the basis of the higher factor loading of each variable either with the first or second principal component. To reflect the proportion of the total variance explained by the first and second component, the factor loadings were multiplied by the proportion of variance explained (for more details see OECD, 2008). The weights for each sub-indicator are summarized in Table 1. The weights were corrected therefore the sum equals one.

Table 1 Standardized eigenvectors and weights

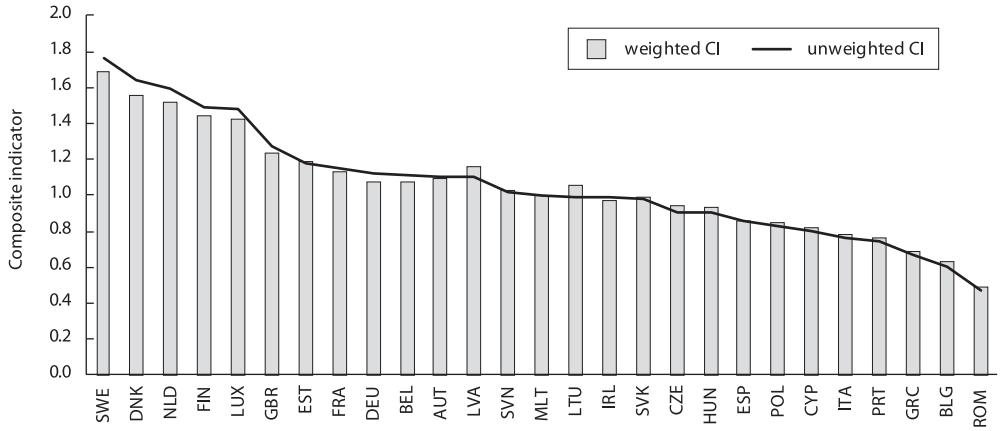
Variable	Prin1	Prin2	Weight
Having computer	0.93	–0.08	0.08
Mobile internet	0.75	0.26	0.05
Phone calls over Internet	0.06	0.78	0.06
Mobile phone satisfaction index	–0.09	0.73	0.05
Internet satisfaction index	–0.23	0.76	0.06
Broadband penetration rate	0.90	–0.20	0.08
E-government usage	0.95	0.02	0.09
Ordering goods over Internet	0.87	–0.28	0.07
Never used the Internet	–0.95	–0.03	0.09
Frequently using the Internet	0.98	0.00	0.09
Using internet banking	0.95	0.09	0.09
High computer skills	0.81	0.22	0.06
High Internet skills	0.38	0.67	0.04
Households with Internet	0.95	–0.08	0.09

Source: Own construction

The highest weights can mostly be seen for the indicators reflecting computer and Internet availability and usage. These variables are then more important for the reflection of differences among the EU states.

Figure 4 depicts the position of the EU states on the basis of the composite indicator, both the weighted and unweighted forms.

Figure 4 Level of e-communication on the basis of composite indicator, EU-27, 2010/2011

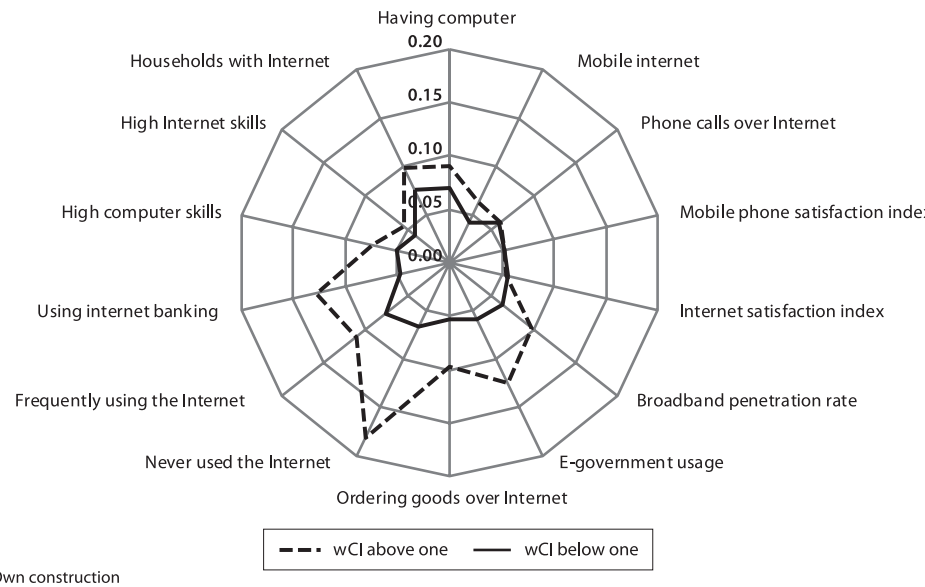


Source: Own construction

As it is seen from the results depicted in Figure 4 the weighted composite indicator reflects better the differences among the member states. This fact can be illustrated by comparing the variability using the coefficient of variation (V). The variation among the EU states is higher in case of weighted composite indicator (V = 30.27%), while in case of unweighted form V = 27.55%.

Figure 5 illustrates the contribution of each sub-indicator to the composite indicator and it enables to identify the sub-indicators which are of the highest differences among the EU. For this purpose the states were split up into two groups. The first group consists of states with the value of wCI above one. The second group of states represents states with the average proportion to median below one. Each point represents the average value for each normalised indicator, separately for the two groups of the states.

Figure 5 Contribution of each sub-indicator to the composite indicator



Source: Own construction

The biggest differences can be recognised from the view of sub-indicators percentage of people never used Internet, percentage of people ordering goods over Internet, egovernment usage and broadband penetration rate.

CONCLUSION

The field of e-communication has been growing rapidly in the last few years, but the level of computer literacy in various countries of the European Union is still unbalanced. It means that the usage of information technology communication is on different levels, too. Western countries show an obvious advantage compared to Eastern ones. This is primarily due to the fact that in these countries, computer training began considerably earlier than in Eastern Europe (see EDCL, 2012). Berry (2011) points out those differences between the countries can be also caused by the different approaches of governments to the support and extended access of the older population and excluded social groups to the ICT.

The use of the information and communication technologies was evaluated by the principal component analysis and by the composite indicators. From the component score plot it is obvious that Romania, Bulgaria, and Greece are at the extreme left, with the lowest level of ICT usage. On the other hand, the Northern states tend to be at the right side of the plot with a high overall ICT usage. The Central European countries tend to be in the upper part of the plot, with a higher-than-average ratio of the quality of services and average ratio of ICT usage. On the other side the coastal states tend to be at the bottom of the plot with a low ratio of the quality of services.

In terms of composite indicators the highest level of e-communication was achieved in the Nordic countries. Sweden occupied the first place, followed by Denmark and the Netherlands. The Czech Republic was 18th from the EU-27 countries. Slovakia reached the best place of the Central European countries on the 17th place, followed by Czech Republic and Hungary on the 19th place.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support from the Faculty of Economic and Management, Czech University of Life Sciences, via IGA grant, No. 20121049, "Analysis of the Internet and computer literacy using data mining techniques".

References

AMERICAN LIBRARY ASSOCIATION. *Presidential Committee on Information Literacy. Final Report* [online]. Chicago: American Library Association, 1989. [cit. 24.6.2012]. <http://www.ala.org/acrl/publications/whitepapers/presidential>.

BERRY, R. *Older People and the Internet; Towards a "System Map" of Digital Exclusion* [online]. 2011. [cit. 24.6.2012]. <www.ilcuk.org.uk>.

DOHNÁLKOVÁ, E., LANDOVÁ, H. *Infogram: Nová platforma pro podporu informačního vzdělávání* (Infogram: A New Platform to Promote Information Literacy). INFORUM 2009: 15th Conference on Professional Information Resources, Prague, 2009.

DOMBROVSKÁ, M. *Information Education at Universities in the Czech Republic – the Overview and the Outlook.* [online]. 2011. [cit. 24.2.2012]. <http://pro.inflow.cz /informacni-vzdelavani-na-vysokych-skolach-v-ceske-republice-–-prehled-vyvoj-dalsi-kroky>.

EBERT, U., WELSCH, H. Meaningful Environmental Indices: a Social Choice Approach. *Journal of Environmental Economics and Management,* 2004, Vol. 47, pp. 270–283.

EDCL. *European Computer Driving License* [online]. 2012. [cit. 24.2.2012]. <http://www.ecdl.cz>.

EUROPEAN COMMISSION. *Eurobarometer 75.1. E-Communication in the Household: Mobile Phone, Television and Internet.* [online]. 2011. [cit. 24.6.2012]. <http://zacat.gesis.org>.

EUROPEAN COMMISSION. *Eurobarometer 75.1: February–March 2011, Study Documentation.* [online]. 2012. [cit. 24.6.2012]. <http://zacat.gesis.org>.

EUROSTAT. *Information Society* [online]. 2011. [cit. 21.8.2012]. <http://epp.eurostat. ec.europa.eu/portal/page/portal/information_society/introduction>.

Hebák, P. et al. *Vícerozměrné statistické metody [3]* (Multivariate Statistical Methods). Prague: Informatorium, 2007.

MONTAGNIER, P., WIRTHMANN, A. Digital Divide: From Computer Access to Online Activities – A Micro Data Analysis. *OECD Digital Economy Papers,* 2011, Vol. 189, p. 73.
MUNDA, G., NARDO, M. *Constructing Consistent Composite Indicators: the Issue of Weights,* EUR 21834 EN, Joint Research Centre, Ispra, 2005.
OECD. *Handbook on Constructing Composite Indicators Methodology and User Guide.* Paris: OECD Publishing, 2008.
RENCHER, A. *Methods of Multivariate Analysis.* 2nd Ed. New York: Wiley Publishing, 2002.
SAISANA, M. et al. *State of the Art Report on Composite Indicators for the knowledge-based Economy.* [online]. 2005. [cit. 28.6.2012]. <http://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/KEI-WP5-D5.1.pdf>.
SAS DOCUMENTATION. SAS / STAT 9.2 User's Guide. Cary, NC: SAS Institute Inc., 2008.

ANNEX

Table A1 Input variables expressed as percentage of population with given characteristic

State	Having computer	Mobile internet	Phone calls over Internet	Mobile phone satisfaction index	Internet satisfaction index	Broadband penetration rate	E-government usage	Ordering goods over Internet	Never used the Internet	Frequently using the Internet	Using internet banking	High computer skills	High Internet skills	Households with Internet
Austria	62	43	32	83	83	24	39	56	18	59	38	42	9	75
Belgium	71	23	28	72	81	30	32	53	14	65	51	28	10	77
Bulgaria	47	12	51	76	78	14	15	13	46	37	2	11	9	45
Cyprus	62	25	42	81	76	23	22	36	41	45	17	23	8	57
Czech Republic	56	47	49	84	85	20	17	41	24	41	23	25	12	67
Denmark	88	50	35	75	77	38	72	77	7	78	71	39	15	90
Estonia	69	43	45	76	80	26	48	27	20	59	65	32	21	71
Finland	79	48	33	73	66	29	58	69	9	76	76	43	19	84
France	75	39	35	70	65	32	36	67	18	62	50	29	13	76
Germany	69	32	21	81	81	31	37	77	16	63	43	25	5	83
Greece	55	24	24	73	81	19	13	33	45	37	6	24	8	50
Hungary	54	26	41	87	85	20	28	32	28	56	19	32	15	65
Ireland	69	42	32	73	72	23	27	56	21	55	34	26	7	78
Italy	61	21	24	79	75	21	17	27	39	49	18	25	12	62
Latvia	69	47	65	79	81	19	31	27	27	53	47	29	31	64
Lithuania	60	36	65	77	84	20	22	25	33	48	37	32	27	62
Luxembourg	78	43	43	73	71	33	55	71	8	76	56	43	13	91
Malta	69	21	35	82	80	29	28	65	30	55	38	24	13	75
Netherlands	95	38	30	74	70	38	59	74	7	79	77	32	19	94
Poland	64	37	34	75	83	15	21	46	33	45	25	18	10	67
Portugal	56	28	12	75	79	20	23	31	41	42	19	28	10	58
Romania	53	15	11	67	65	14	7	13	54	24	3	10	7	47
Slovakia	55	37	45	71	78	16	35	48	20	56	33	23	12	71
Slovenia	75	51	22	77	83	24	40	45	29	54	29	31	16	73
Spain	59	29	13	70	71	23	32	39	29	48	27	32	11	64
Sweden	91	59	32	79	83	32	62	75	5	80	75	42	20	91
United Kingdom	73	52	30	66	63	31	40	82	11	70	45	32	11	83

Source: Own construction, Eurobarometer 75.1, Eurostat

Gender and Extended Actuarial Functions in Pension Insurance

Jana Špirková[1] | *Matej Bel University, Banská Bystrica, Slovakia*
Mária Spišiaková | *Matej Bel University, Banská Bystrica, Slovakia*

Abstract

This paper brings analysis of the impact of a ban on the use of gender in insurance, with special stress on pension annuity, according to the requirements of the European Court of Justice. The paper brings a state-of-the-art overview of known and extended actuarial functions which relate to modeling of a premium of endowment, term life insurance and pension annuity. Moreover, the amounts of the pension annuities payable thly per year in a model of the third pillar pension are modeled and analyzed for different interest rates using life tables for both genders and unisex.

Keywords

Annuity, extension, gender, premium, pension

JEL code

C 13, G 22

INTRODUCTION

The European Union (EU) Gender Directive (Council Directive 2004/113/EC) guarantees equal treatment between men and women in the access to and supply of goods and services. However, the Directive does not prohibit insurers from using gender in the calculation of premiums and benefits, as it contains an exemption to this rule: under Article 5(2), Member States can opt out from banning the use of gender and can allow proportionate differences in insurance premiums and benefits where the use of gender is a determining factor in the assessment of risk based on the relevant and accurate actuarial and statistical data, provided that Member States ensure that such data is compiled, published and regularly updated. All European national legislative assemblies approved the option to use the opt-out for life products – including life insurance and pension annuities. However, on 1st March 2011, the European Court of Justice (ECJ) ruled that this time-unlimited opt-out provision from the EU Gender Directive was inconsistent with the European Charter (Test-Achats ruling). The ECJ ruled that the (time-unlimited) exemption is invalid but allowed for a transition period for implementation up to 21st December 2012.

[1] Faculty of Economics, Matej Bel University, Tajovského 10, 975 90 Banská Bystrica, Slovakia. Corresponding author: e-mail: jana.spirkova@umb.sk, phone: (+421)484466626.

National governments of Member States will be obliged to change their laws accordingly by this date. For more information see also Oxera (2012).

Not only the risk factor – tender, but also actuarial modeling can affect the size of the premium and future size of premiums and pension annuities. The net annual premium of whole life annuities and also other actuarial functions are usually evaluated for integer ages and terms, assuming that cash flows are payable annually. However, annuities are very often paid more frequently than annually, namely monthly, quarterly, but also semi-yearly. We can find in various sources well known formulas for valuation of certain and expected annuities which are paid more frequently than annually, see Booth et al. (1999), Dickson (2009), MacDonald (2012). However, it is possible to extend these actuarial functions, which means to determine them much more accurately.

We introduce much more precise formulas of the expected present value of the annuities payable m times a year on the basis of Woolhouse's formula regarding Maclaurin expansion in MacDonald (2012). With respect to the range of the paper, we only introduce an outline of the derivation. The whole derivation can be found in Špirková, Urbaníková (2012). Moreover, we apply extended actuarial functions on the evaluation of monthly pension annuities in a model of the third pillar pension. Synchronously, we compare the size of monthly pension annuities with respect to life tables of male, female and unisex.

This paper is organized as follows: in the first part we recall two factors which impact the size of premium and also the size of future pension annuities – EU Gender Directive and mathematical model of the expected present values and accumulated future values of annuities payable annually and m-thly per year. We present Woolhouse's formula, which was derived according to Maclaurin expansion, see MacDonald (2012). In the second part we develop formulas for m-thly annuities payable in advance by the mentioned Woolhouse's formula. In the third part we apply the above-mentioned developed formulas for m-thly paid annuities, especially the expected present value of m-thly annuity payable in advance, and the accumulated future value of m-thly certain annuity payable in advance on the determination of monthly pension annuities in a simplified model of the third pillar pension. At the end, we give some remarks and schemes of our next investigation.

1 PRELIMINARIES

Note that in the case of positive cash-flows the present value with payments paid m-thly in advance is less than the corresponding yearly present value. In the yearly case, a full payment of the annuity of one monetary unit would be made at time 0. In the m-thly case, the annuitant receives $1/m$ of the monetary unit each period of the length $1/m$, and will not have collected the full one amount until one m-th of a year before the year end. This is true for each year of the annuity. There is therefore a loss of interest, which is reflected in a lower present value for the m-thly case.

Exact calculations of Euler-Maclaurin expansion are not always used in practice. It is common to use approximations. A one-year, one-unit annuity payable m-thly is simply an annuity paying $1/m$ units for m periods. We will refer to these as m-thly annuities or annuities payable m-thly. In this way we can derive the corresponding results for the expected present value of an m-thly paid annuity. We consider here a whole annuity-due of one monetary unit per annum, payable to a life age x with $1/m$ payable at the beginning of each m-thly period. Consider a whole life annuity with payments of $1/m$ made m times a year at moments $0,1/m,\ldots,(m-1)/m$. So, for each complete year, the annual premium is one. Say, if $m = 12$, an m-th is a month.

Proposition 1 uses Woolhouse's formula for annuities in advance which can be derived according to Maclaurin expansion. Regarding the range of our paper, we do not provide the whole derivation of Woolhouse's formula, but it can be found for example in Dickson (2009).

Proposition 1 – Woolhouse's formula
Let *f(x)* be a continuous function on an interval [*a, b*]where $h = \frac{b-a}{N}$ is the width of individual intervals and *N* is the number of divisions of the interval [*a, b*], *m* is the number of divisions of the interval of the width *h*, 1<*m*, $m \in Z$.
Then the approximation:

$$\frac{1}{m}f(a) + \frac{1}{m}f\left(a + \frac{h}{m}\right) + \cdots + \frac{1}{m}f\left(a + \frac{(Nm-1)h}{m}\right) \approx$$
$$\approx f(a) + f(a+h) + \cdots + f(a+(N-1)h) - \frac{m-1}{2m}\left(f(a) - f(a+Nh)\right) + \quad (1)$$
$$+ \frac{h}{12} \times \frac{m^2-1}{m^2}\left(f'(a) - f'(a+Nh)\right) - \frac{h^3}{720} \times \frac{m^4-1}{m^4}\left(f'''(a) - f'''(a+Nh)\right)$$

holds.

If we want to derive the present value of *m*-thly paid annuities in the size of 1/*m* of monetary unit according to (1) the function *f* is represented by $f(t) = {}_tp_x \times v^t$ as a function of time *t*, where ${}_tp_x$ is the probability that *x*-aged man will be alive at age $x + t$ and $v = 1/(1 + i)$ is a discounting factor, *i* is a technical interest rate. If we do not assume terms with derivatives, then the value *f*(*a*) at (1) for $t = 0$ is 1, and the value $f(a + Nh)$ for $t \to \infty$ tends to 0. Seeing that the left side of equation (1) represents the expected present value of the whole annuity of one per year, payable to the entry aged *x* with 1/*m* at the beginning of each *m*-thly period, we get the approximation:

$$\ddot{a}_x^{(m)} \approx \ddot{a}_x - \frac{m-1}{2m}, \quad (2)$$

where $\ddot{a}_x$ is the present value of whole life yearly annuities in the amount of 1 monetary unit, $\ddot{a}_x = \sum_{t=0}^{\omega-x} {}_tp_x \times v^t$ or by the commutation functions $\ddot{a}_x = \frac{N_x}{D_x}$. Commutation function $D_x = l_x \times v^x$ represents the discounted number of survivors at age *x*, and $N_x = \sum_{i=0}^{\omega-x} D_{x+i}$.

The theory about commutation functions can be found, for example, in Gerber (1997), Urbaníková, Vaculíková (2006).

2 EXTENDED FORMULAS OF ANNUITIES

If we would like to extend the approximation (2) with at least one term with derivative from (1), we can do the next reflection: we assume the force of mortality μ_x at time *t* constant and then the probability that *x*-age insured group will be alive at age $x + t$ is given by:

$${}_tp_x = e^{-\mu_x t}. \quad (3)$$

The value μ_x is known as the force of mortality at age *x* and has several equivalent forms, see also Promislow (2006), Gerber (1997). The quantity μ_x gives us the "relative rate" of decline in this group at age *x*.

For example, from (3) we can see that the force of mortality at age *x* can be rewritten by:

$$\mu_x = -\frac{\frac{d_tp_x}{dt}}{{}_tp_x}. \quad (4)$$

We can view μ_x as the force of mortality at time *t* for an individual age *x* and approximate by $\mu_x \approx -\ln p_x$.

The value $\delta = \ln(1 + i)$ represents the so-called force of interest. From the derivative of the function ${}_tp_x \times v^t$ according to *t* we get the expression below:

$$\frac{d}{dt}\,{}_tp_x \times v^t = -(\mu_x + \delta) \times {}_tp_x \times v^t. \tag{5}$$

Regarding the previous consideration we get a more precise approximation of the expression (2), which is given by:

$$\ddot{a}_x^{(m)*} \approx \ddot{a}_x - \frac{m-1}{2m} - \frac{m^2-1}{12m^2} \times (\mu_x + \delta). \tag{6}$$

Although the values $\mu_{x,}$ δ are independent, for $t \to 0$ we can approximate the mentioned sum by $2i$ (according to their average values and also for simplicity), we can rewrite the previous formula by:

$$\ddot{a}_x^{(m)**} \approx \ddot{a}_x - \frac{m-1}{2m} - \frac{m^2-1}{6m^2} \times i\,. \tag{7}$$

Similarly, we can derive an expression for the present value of deferred whole life annuity in advance ${}_{k|}\ddot{a}_x^{(m)}$ which is paid m-thly per year in the amount $1/m$ of monetary unit from age $x + k$. It is given by:

$${}_{k|}\ddot{a}_x^{(m)} \approx {}_{k|}\ddot{a} - \frac{m-1}{2m}\frac{D_{x+k}}{D_x}\,. \tag{8}$$

An extended version of the previous formula is as follows:

$${}_{k|}\ddot{a}_x^{(m)*} \approx {}_{k|}\ddot{a}_x - \frac{D_{x+k}}{D_x} \times \left(\frac{m-1}{2m} + \frac{m^2-1}{12m^2} \times (\mu_x + \delta)\right), \tag{9}$$

or in a simplified form:

$${}_{k|}\ddot{a}_x^{(m)**} \approx {}_{k|}\ddot{a}_x - \frac{D_{x+k}}{D_x} \times \left(\frac{m-1}{2m} + \frac{m^2-1}{6m^2} \times i\right). \tag{10}$$

For the purposes of our model we recall the accumulated value of temporary m-thly paid certain annuities of $1/m$ monetary unit which are payable in advance during n years, which is denoted as $\ddot{s}_{\overline{n}|}^{(m)}$ and is as follows:

$$\ddot{s}_{\overline{n}|}^{(m)} = \frac{1}{m} \times q^{\frac{1}{m}} \times \frac{q^n - 1}{q^{\frac{1}{m}} - 1}. \tag{11}$$

For more information see also Sekerová, Bilíková (2007), Urbaníková (2008).

Remark 1

According to Woolhouse`s formula we can derive approximation of formula (11) according to $f(t) = (1 + i)^{n-t}$. Our model represents a deterministic approach in pension annuities. Stochastic models are described, for example in Booth et al. (1999), Gerber (1997), Potocký et al. (2007), Potocký (2008).

3 APPLICATION OF EXTENDED ACTUARIAL FUNCTIONS

In the previous part we introduced the extension of some basic formulas for expected life annuities. Now, we will apply the mentioned formulas for determination and analysis of pension annuities from a simplified model of the third pillar pension, which represents a whole life annuity payable monthly in advance. Of course, we know that many factors such as raising of wage, inflation, costs of Pension Asset Management Companies, costs of insurance companies, etc., will influence the future pension. But we would like to emphasize that in our model we can show how the risk factor gender and extended actuarial functions, namely the present value of m-thly paid annuities $\ddot{a}_x^{(m)}$, and its

extensions $\ddot{a}_x^{(m)**}$ or $\ddot{a}_x^{(m)*}$, influence the size of future pension annuities. In our model we evaluate the accumulated value of certain annuities in advance (11) as future accumulated value of funds, from which pension annuities will be paid out. We introduce basic formulas for determination of monthly annuities from the third pillar pension.

With regard to (2) and (11), monthly pension annuities can be determined as follows:

$$p^{(m)} = \frac{A \times \ddot{s}_{\overline{n}|}^{(m)} \times (1-p)}{m \times \ddot{a}_x^{(m)}}, \tag{12}$$

where A is a monthly payment of annuities during n years, $A \times \ddot{s}_{\overline{n}|}^{(m)}$ represents the real accumulated value over duration time n, p represents the first higher pension (in percent).

For more precise evaluation we can apply a precise formula according to (6), (11) which is given by:

$$p^{(m)*} = \frac{A \times \ddot{s}_{\overline{n}|}^{(m)} \times (1-p)}{m \times \ddot{a}_x^{(m)*}}. \tag{13}$$

We can express the previous formula according to (7) and (11) as follows:

$$p^{(m)**} = \frac{A \times \ddot{s}_{\overline{n}|}^{(m)} \times (1-p)}{m \times \ddot{a}_x^{(m)**}}. \tag{14}$$

Remark 2

m-thly paid pension annuities can be modified according to the requirements of clients. For example, they can be expressed as follows:

$$p^{(m)\circ} = \frac{A \times \ddot{s}_{\overline{n}|}^{(m)} \times (1-p)}{m \times (\ddot{a}^{(m)*} + z \times \ddot{a}_{\overline{n}|}^{(m)} A_x)}, \tag{15}$$

where z is the coefficient of the survivor´s pension arrangements, $\ddot{a}_{\overline{n}|}^{(m)}$ represents the present value of temporary m-thly paid certain annuities of $1/m$ monetrary unit which are payable in advance during n years and is given by $\ddot{a}_{\overline{n}|}^{(m)} = \frac{1}{m} \times q^{\frac{1}{m}} \times \frac{1-q^{-n}}{q^{\frac{1}{m}}-1}$, the value A_x is the net single premium of whole life insurance for x-aged client with the policy value of one monetary unit. It can be expressed by commutation functions as follows $A_x = \frac{M_x}{D_x}$, where $M_x = \sum_{i=0}^{\omega - x} C_{x+i}$ and $C_x = d_x \times v^{x+\frac{1}{2}}$ is the discounted number of deaths at age x. The whole evaluation is based on life tables from Mortality tables (2012).

In our model we assume that a client at the age of 18 saves monthly 30 euros in a pension company during the whole duration time until retirement. The first higher pension is 10% from accumulated value.

Firstly, we recall that *m*-thly pensions according to formulas (13)–(15) are not significantly different for small technical interest rates. For smaller technical interest rates around the recent interest rate i = 2.5% p.a. the difference between using known and extended formulas for pension from the third pillar pension is only a few cents. If we assume a technical interest rate i = 7.5% p.a., the values differ in a few euros.

In Table 1 are evaluated monthly annuities from our model with entry age to retirement.

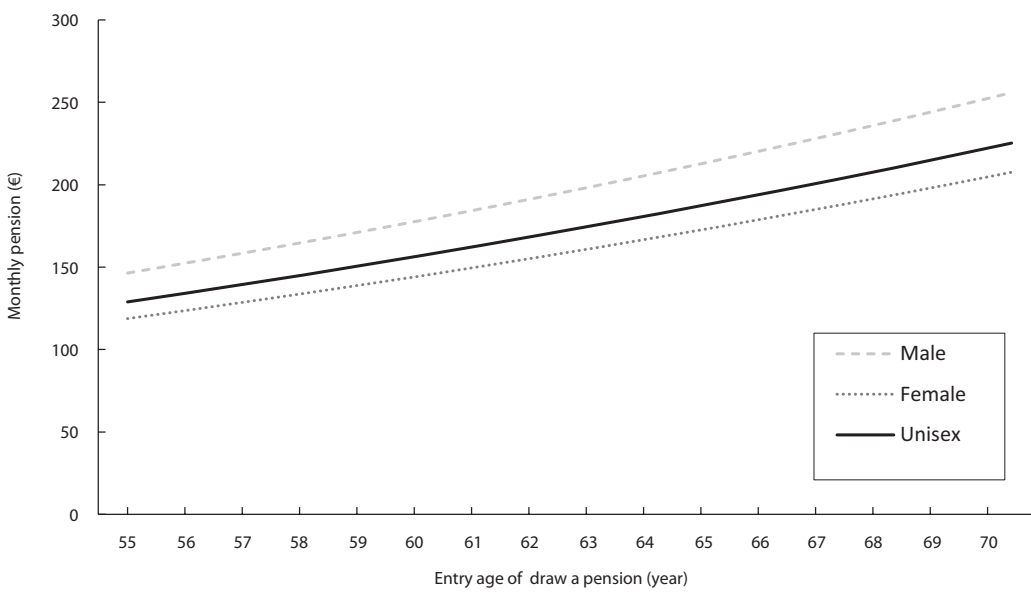
Table 1 Monthly annuities (in euros) for client from our model according to formula (12)

	i = 2.5% p.a.			*i* = 7.5% p.a.		
x	**male**	**female**	**unisex**	**male**	**female**	**unisex**
55	146.31	118.76	128.83	639.16	544.41	580.70
56	152.46	123.72	134.20	690.64	588.26	627.48
57	158.68	128.80	139.72	745.98	635.40	677.76
58	165.09	134.01	145.36	805.48	686.07	731.81
59	171.67	139.35	151.16	869.43	740.55	789.91
60	178.41	144.82	157.09	938.18	799.11	852.38
61	185.32	150.43	163.17	1 012.09	862.06	919.53
62	192.40	156.18	169.41	1 091.54	929.73	991.71
63	199.66	162.07	175.80	1 176.95	1 002.48	1 069.31
64	207.10	168.11	182.35	1 268.77	1 080.68	1 152.73
65	214.73	174.30	189.07	1 367.47	1 164.76	1 242.40
66	222.55	180.65	195.95	1 473.57	1 255.13	1 338.80
67	230.56	187.15	203.01	1 587.64	1 352.28	1 442.43
68	238.77	193.82	210.24	1 710.25	1 456.72	1 553.84
69	247.19	200.65	217.65	1 842.07	1 569.00	1 673.60
70	255.82	207.66	225.25	1 983.77	1 689.69	1 802.33

Source: Own construction

In Figure 1 we can see the dependence of the size of monthly pension according to entry age separately for male, female and unisex according to (13) and corresponding to life tables, which were determined with respect to a technical interest rate $i = 2.5\%$ p.a. Based on the data available for the Slovak Republic with the technical interest rate 2.5% p.a. men (aged 55 and more) could see a reduction in pension income from pension annuities of around 12% or more on average; women (aged 55 and more) could see pension income rise of around 8% or more on average, see Table 1. Moreover, additional costs could arise from insurers applying a gender mix risk premium due to the risk of adverse selection. There could also be additional marketing costs due to a ban on the use of gender.

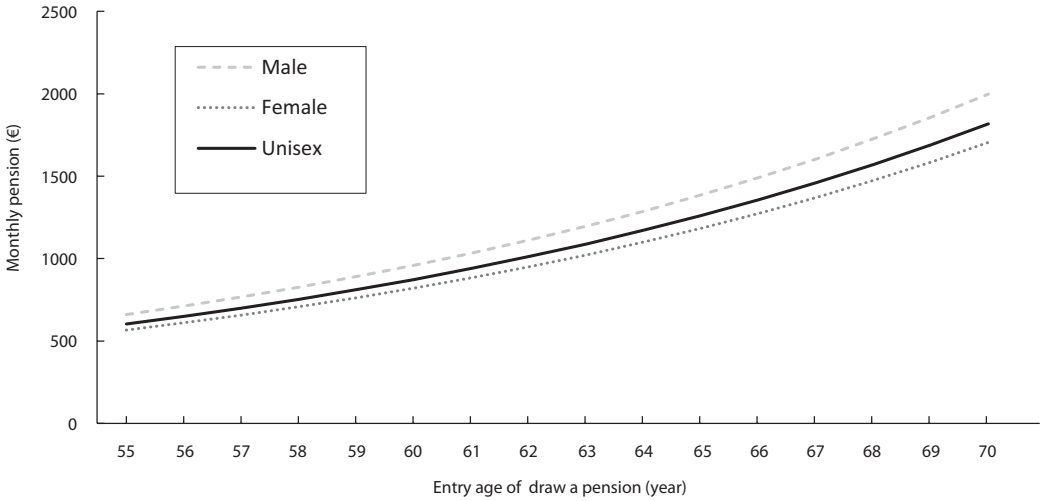
Figure 1 Dependence of the size of monthly pension according to entry age (technical interest rate 2.5% p.a.)



Source: Own construction

With a technical interest rate 7.5% p.a. men (aged 55 and more) could see a reduction in pension income from pension annuities of around 9% or more on average; women (aged 55 and more) could see pension income rise of around 7% or more on average, see Table 1 and Figure 2.

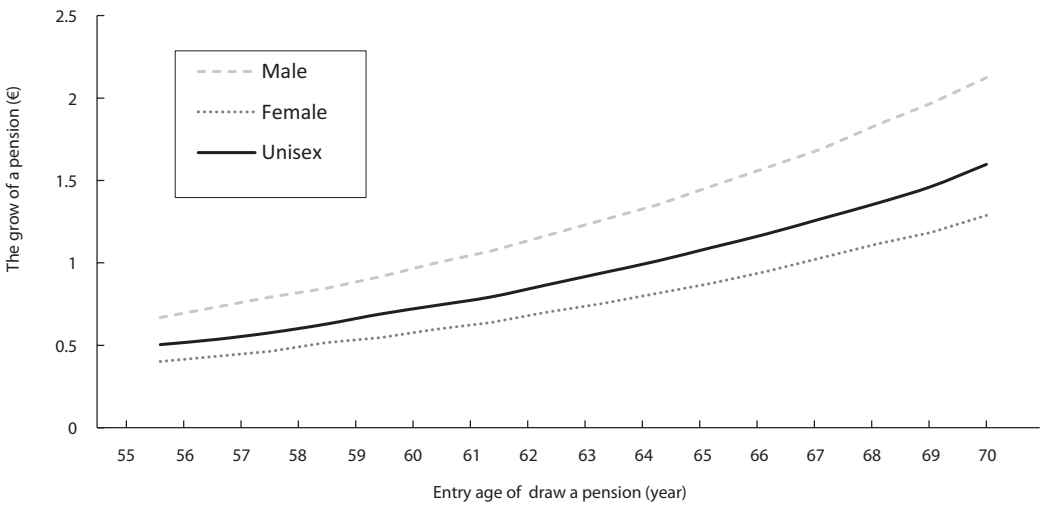
Figure 2 Dependence of the size of monthly pension according to entry age (technical interest rate 7.5% p.a.)



Source: Own construction

In Figure 3 is shown the difference in the size of pension annuity based on the comparison of formulas (12) and (13) regarding the interest rate 7.5% p.a.

Figure 3 The difference in the size of the pension with respect to formulas (12) and (13) according to entry age (technical interest rate 7.5% p.a.)



Source: Own construction

All values and graphs were evaluated and constructed by MS Office Excel 2010 system.

CONCLUSION

In our paper we discuss the size of pension annuities with respect to the rule of ECJ and well-known and extended formulas for -thly paid annuities.

For annuities, women receive a lower pension annuity payment monthly than men for the same accumulated value in the time of retirement. However, women have a higher life expectancy, which means that women receive pension annuities over a longer time, and so women receive the same expected lifetime annuity benefit as men. Acording to Oxera (2012) focusing on the analysis of the gender as a risk factor in insurance in selected European countries, namely Germany, France, Spain, Poland, Czech Republic and Belgium, men could see a reduction in pension income from pension annuities of around 5% or more on average. However, this source does not study the dependence on the interest rate.

Based on the data available for the Slovak Republic with the technical interest rate 2.5% p.a. men could see a reduction in pension income from pension annuities of around 12% or more on average and women could see a pension income rise of around 8% or more on average. With a technical interest rate 7.5% p.a. a reduction for men is around 9% or more and a pension income rise of around 7% or more on average for women.

From a mathematical point of view it is interesting to extend the mentioned formulas. But their application is useful only for a very long duration of time and higher monthly payments and higher technical interest rates. Moreover, the mentioned extended formulas are suitable for modeling of second and third pillar pensions and also deferred pension annuities.

ACKNOWLEDGEMENTS

This work was supported by the Civic Association ECONOMY.

References

BOOTH, P. et al. *Modern Actuarial Theory and Practice.* Chapman & Hall / CRC, 1999.

DICKSON, D. C. M. et al. *Actuarial Mathematics for Life Contingent Risks.* New York: Cambridge University Press, 2009.

GERBER, H. U. et al. *Life Insurance Mathematics.* Berlin, Heidelberg, New York : Springer-Verlag, 1997.

MACDONALD, A. S. *Euler-Maclaurine Expansion and Woolhouse's Formula* [online]. [cit. 3.6.2011]. <http://perso.univ-rennes1.fr/arthur.charpentier/euleract.pdf>.

MORTALITY TABLES. *Mortality tables* [online]. [cit. 10.3.2012]. <http://portal.statistics.sk/showdoc.do?docid=2464>.

OXERA. *The Impact of a Ban on the Use of Gender in Insurance* [online]. [cit. 14.3.2012]. <http://www.slaspo.sk/tmp/asset_cache/link/0000033890/11207%20oxera-study-on-gender-use-in-insurance.pdf>.

POTOCKÝ, R. On a Dividend Strategy of Insurance Companies. *Ekonomie a Management*, Vol. 4, 2008, pp. 103–109.

POTOCKÝ, R., STEHLÍK, M. *Stochastic Models in Insurance and Finance with Respect to Basel II. JAMSI.* 2007, No. 2, Vol. 3, pp. 237–245. ISSN 1336-9180.

PROMISLOW, D. *Fundamentals of Actuarial Mathematics.* Toronto, Canada: John Wiley & Sons, 2006.

SEKEROVÁ, V., BILÍKOVÁ, M. *Insurance Mathematics* (in Slovak). Bratislava: Ekonóm, 2007.

ŠPIRKOVÁ, J., URBANÍKOVÁ, M. *Actuarial Mathematics – Life Insurance* (in Slovak). Bratislava: Iura Edition, spol s r.o., 2012.

URBANÍKOVÁ, M. *Financial Mathematics* (in Slovak). Nitra: Constantine The Philosopher University, 2008.

URBANÍKOVÁ, M.,VACULÍKOVÁ, Ľ. *Actuarial Mathematics* (in Slovak). Bratislava: Slovak University of Technology, 2006.

LOGIT and PROBIT Models in the Probability Analysis: Change in the Probability of Death of Celiac Disease Patients

Ondřej Šimpach[1] | *University of Economics, Prague, Czech Republic*

Abstract

It is estimated, that in the Czech Republic live about 40 000–50 000 people suffering from celiac disease, which is a disease of gluten intolerance. At the beginning of the independent Czech Republic, the life expectancy at birth of these people was quite low, because just in this period detailed diagnosis of this disease came from abroad. With an increasing age the probability of death of these people grew faster than that of total population. The aim of this study is to analyse the probability of death of x-year old persons during next five years after the general medical examination in 1990 and 1995. Both analyses will be solved using LOGIT and PROBIT models and the hypothesis claiming, that probability of death of x-year old person suffering from celiac disease decreased few years after the gaining of new medical knowledge from abroad will be confirmed or refused.

Keywords

Probability of death, celiac disease, LOGIT, PROBIT, discrete dependent variables

JEL code

C35, C40, I19

INTRODUCTION

Medicine provides to people new knowledge about the diagnosis of specific diseases. By the end of the past regime in the Czech Republic, a detailed diagnosis of disease of the small intestine mucosa, professionally known as celiac disease (see Společnost pro bezlepkovou dietu / Society for the gluten-free diet) was not known. At present, it is estimated, that in the Czech Republic there are about 40 000–50 000 inhabitants, who suffer from this disease. Unfortunately, only about 10–15% of all patients are under medical supervision (see Poradenské centrum pro celiakii a bezlepkovou dietu / Advice center for celiac disease and gluten-free diet). In the past, the people suffering from this disease did not have the sufficient information about correct nutrition and thus they could not eliminate the consequences caused by

[1] University of Economics, Prague, Winston Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: ondrej.simpach@vse.cz, phone: (+420)224095273, (+420)737665461.

normal diet. Normal diet, which also includes food containing gluten, etched the small intestinal mucosa of these people and most of them had not been able to live as long ordinary people (see e.g. Logan et al., 1989) did. It is important to note, that we do not have separate statistics of people who died of this disease. Deaths are recorded in the summary group "XI - Diseases of the Digestive System", published by the Czech Statistical Office (CZSO).

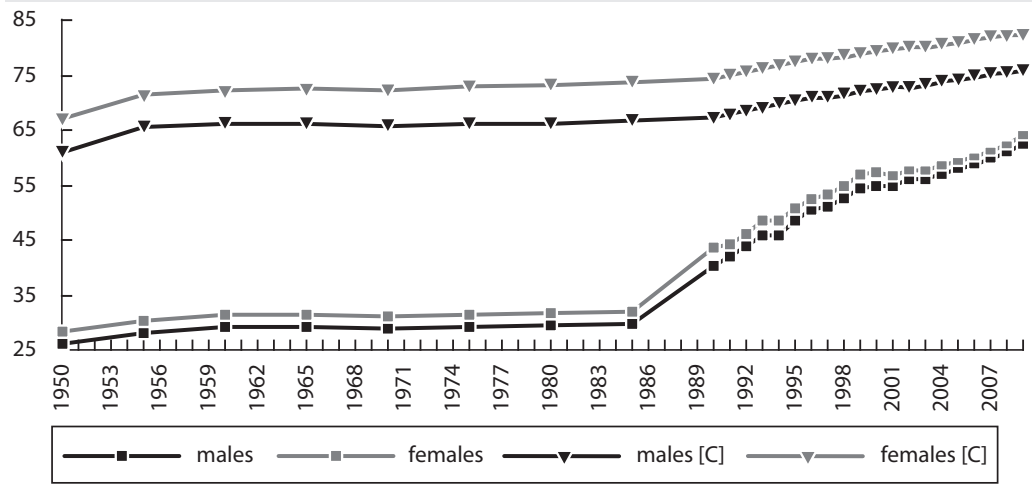
In the Czech Republic, there live about tens of thousands of people who suffer from other diseases. But the diagnoses of these diseases are known for a long time. Taking the developed drugs eliminates partly or completely the consequences of these diseases. This study will make specific probabilistic analysis of death of x-year old individuals. Using the LOGIT and PROBIT models the probabilities of death of x-year old persons suffering from celiac disease during next 5 years after the general medical examination in 1990 and the probabilities of death of x-year old persons suffering from any of other diseases during next 5 years after the general medical examination will be calculated in 1990. Diagnoses of these other diseases were known for longer time than celiac. These are mainly the following:

- Diabetes,
- High blood cholesterol (greater than 5 mmol / l),
- High blood pressure (more than 150 SYS and 95 DIA),
- Bronchial asthma,

  or other diseases.

These diseases were selected because they are frequently represented in the population of the Czech Republic. The probabilities of death will be calculated again for both groups (the group of people suffering from celiac disease and the group of people suffering from any other disease), but the data will be shifted by 5 years forward in time, (i.e. taken in 1995). Based on the results from these other estimates, the hypothesis claiming, that the probabilities of death of x-year old persons suffering from celiac disease decreased in time will be acknowledged. The evolution of life expectancy at birth for males and females suffering from celiac disease is shown in Figure 1. The time series labelled "males" and "females" show the actual life expectancy at birth of total population, which is published by the CZSO. Time series "males_c" and "females_c" show the estimated life expectancy of people suffering from celiac disease. Values for the years 1950–1990 are given only in the 5-year time points, because for the purposes of this study they are not necessary.

Figure 1 Estimates of life expectancy at birth of males and females suffering from celiac disease in 1950–2009



Source: Czech Statistical Office, Šimpach (2011)

After the estimation of LOGIT and PROBIT parameters models of probabilities of deaths of x-year old persons will be graphically interpreted.

1 ASSUMPTIONS OF THE STUDY

For the experiment of nonlinear regression, applied in the first part of this study 243 observations of variables consisting of two samples were obtained. It is important to note, that this is not a representative selection for the application of standard methods of mathematical statistics. The selection was not taken at random. This is the only existing data matrix, obtained by own research. The data matrices, which are prepared by health insurance funds, do not have required form and all needed variables are not recorded. The data contains 118 observations of females and 125 observations of males. (This proportion is set by the fact that the proportion of girls at birth is approximately 0.485 and the proportion of boys at birth is approximately 0.515. For more information see e.g. Pavlík et al., 1986). The selection is obtained from 3 doctors (from 3 different parts of the Czech Republic), whose specialization were adult patients suffering from diseases of digestive system and any other diseases. Selection consists all individual invited in 1990 to general medical examination and their health status was checked in the future.

For consecutive experiment of nonlinear regression, applied in the second part of the study, other 245 observations of patients, consisting of two samples were obtained. The data contains 119 observations of females and 126 observations of males. It is a selection of patients invited in 1995 to the overall medical examination and their health status was checked in the future (but obtained from other 3 doctors from other 3 different parts of Republic). It is almost zero probability that some patients from the first sample are contained in the second sample.

The authors Spector and Mazzeo (1980) put together an example, where they estimated the probability with which a student will succeed in the exam. Based on this example probabilistic LOGIT and PROBIT models were created, which are currently used by many authors in their calculations and publications, such as Hoyos et al. (2010) or Yang and Raehsler (2005). For this study LOGIT and PROBIT models were compiled estimating the probabilities of death of x-year old persons during next five years after the general medical examination, if the disease of the digestive system (known as celiac disease now) was diagnosed for an individual and the probabilities of death of x-year old person during next five years after the general medical examination, if any other disease was diagnosed for a respective individual (e.g. diabetes, high blood cholesterol, high blood pressure, bronchial asthma or other).

Dependent variable Y is an alternative. The value of variable Y equals 1 → the person will die, or the value of variable Y equals 0 → the person will survive. In order to determine the values lying between the two extremes, LOGIT and PROBIT models will be applied where the dependent variable takes values from interval <0, 1>.

The variables in selected data matrices are the following:

- *AGE* – is the age of the person, invited to the general medicinal examination.
- *CIRD* – Constant Increased the Risk of Death, which was designed primarily for the purpose of this study and the calculation of this constant gives the Formula (1). The constant takes values from interval <5, 35>.
- *CEL* – is a binary variable, where the value "0" = a person does not suffer from celiac disease, but suffers from any of the other diseases listed above, or "1" = a person suffers from celiac disease.
- *DEATH_5* – is a binary variable, where the value "1" = a person during next 5 years after the general medicinal examination died as a result of the disease (celiac or other), or "0" = a person during next 5 years after the general medicinal examination did not die as a result of the disease (celiac or other).

As stated above, the constant increasing the risk of death was designed only for the purpose of this study and its calculation is based on Table 1.

Table 1 The responses to doctor during the general medical examination

	$Ш_1$	$Ш_2$	$Ш_3$
Smoker	no	occasionally	periodically
Black coffee	no	occasionally	periodically
Alcohol	no	occasionally	periodically
Sleeping	regularly	irregularly	very bad
Eating	regularly	irregularly	very bad

Source: Own construction

The doctor asked the patient few simple questions during the general medicinal examination. Based on the responses, the table below was created, where verbal responses were replaced by "0" and "1", where "0" = patient's response does not coincide with the word specified in the relevant cell and "1" = patient response coincides with the word specified in the relevant cell. In each row of the table can be only one value "1".

Table 2 The responses to doctor during an overall medical examinations in "0/1" format

	$Ш_1$	$Ш_2$	$Ш_3$
Smoker	$Ш_{i,j} = 0/1$	…	…
Black coffee	…		
Alcohol	…		
Sleeping	…		
Eating	…		

Source: Own construction

After recording the responses to Table 2 in the "0/1" format, there was prepared following formula:

$$CIRD = \left(1 \times \sum_{i=1}^{5} Ш_{i,1}\right) + \left(3.5 \times \sum_{i=1}^{5} Ш_{i,1}\right) + \left(7 \times \sum_{i=1}^{5} Ш_{i,3}\right). \quad (1)$$

This Constant can take values from interval <5, 35>, where the extreme value of 5 means, that the patient does not increase the risk of death because of its poor lifestyle and extreme value of 35 means, that the patient increases the risk of death in the worst way possible. Weights 1, 3.5 and 7 were set on the recommendation of the doctors, who provided data matrices. The general rule, arising from the literature, is not here.

Another study, but for case of USA prepared Rubio-Tapia et al. (2009), where they had better data matrices and used other statistical approach. Authors used data matrices including the age of testing and the age at death. Bigger and exhaustive sample allowed for the application of standard methods of mathematical statistics.

1.1 LOGIT model

From the assumptions listed by Christensen (1990) is the probability function:

$$P_i = E(Y = 1|\boldsymbol{X}_i) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta}' \boldsymbol{X}_i)}}, \quad (2)$$

modified for the purpose of this study to form:

$$P_i = E(Y = 1|\boldsymbol{X}_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 AGE + \beta_2 CIRD_i + \beta_3 CEL_i)}}, \quad (3)$$

where i is $(i)^{th}$ patient. Let us denote:

$$Z_i = \beta_0 + \boldsymbol{\beta}' \boldsymbol{X}_i, \quad (4)$$

which is modified for the purpose of this study to form:

$$Z_i = \beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i \tag{5}$$

and following formula:

$$P_i = \frac{1}{1 + e^{-Z_i}} = \frac{e^{Z_i}}{1 + e^{Z_i}} = F(Z_i), \tag{6}$$

is distribution function of the logistic distribution. Probability of death of x-years old person who will not die during next five years after the overall medical examination is:

$$1 - P_i = \frac{1}{1 + e^{Z_i}} \tag{7}$$

and therefore:

$$\frac{P_i}{1 - P_i} = e^{Z_i}. \tag{8}$$

By the logarithm we obtain LOGIT:

$$\text{In}\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_0 + \boldsymbol{\beta' X_i}, \tag{9}$$

which is for the purpose of this study modified to form:

$$\text{In}\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i. \tag{10}$$

To estimate the unknown parameters of the LOGIT model we can't use classical method of least squares, but with quality software we can use the maximum likelihood method. From the common formula of the log-likelihood function, specified e.g. by Christensen (1990):

$$\text{In } L(\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{N} \left[Y_i \text{ In}\left(\frac{e^{Z_i}}{1 + e^{Z_i}}\right) + (1 - Y_i) \text{In}\left(1 - \frac{e^{Z_i}}{1 + e^{Z_i}}\right)\right], \tag{11}$$

after substitution:

$$\text{In } L(\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{N} \left[Y_i \text{ In}\left(\frac{e^{\beta_0 + \boldsymbol{\beta' X_i}}}{1 + e^{\beta_0 + \boldsymbol{\beta' X_i}}}\right) + (1 - Y_i) \text{In}\left(1 - \frac{e^{\beta_0 + \boldsymbol{\beta' X_i}}}{1 + e^{\beta_0 + \boldsymbol{\beta' X_i}}}\right)\right], \tag{12}$$

and for the purpose of this study is:

$$\begin{aligned} \text{In } L(\beta_0, \beta_1, \beta_2, \beta_3) &= \sum_{i=1}^{N} \left[DEATH\_5_i \text{ In}\left(\frac{e^{\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i}}{1 + e^{\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i}}\right)\right. \\ &\left. + (1 - DEATH\_5_i) \text{ In}\left(1 - \frac{e^{\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i}}{1 + e^{\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i}}\right)\right]. \end{aligned} \tag{13}$$

1.2 PROBIT model

To estimate the unknown parameters of the PROBIT model we can not use classical methods of least squares either, but we can use a universal maximum likelihood method. From the common formula of the log- likelihood function, specified e.g. by Christensen (1990):

$$\text{In } L\ (\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{N} [Y_i \text{ In } (F(Z_i)) + (1 - Y_i) \text{ In } (1 - F(Z_i))], \tag{14}$$

after substitution:

$$\text{In } L\ (\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{N} [Y_i \text{ In } (F(\beta_0 + \boldsymbol{\beta' X_i})) + (1 - Y_i) \text{ In } (1 - F(\beta_0 + \boldsymbol{\beta' X_i}))], \tag{15}$$

for the purpose of this study:

$$\begin{aligned} &\text{In } L\ (\beta_0, \beta_1, \beta_2, \beta_3) \\ &\quad = \sum_{i=1}^{N} [DEATH\_5_i \text{ In } (F(\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i)) \\ &\quad + (1 - DEATH\_5_i) \text{ In } (1 - F(\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i))] \end{aligned} \tag{16}$$

and distribution function:

$$F(\beta_0 + \boldsymbol{\beta' X_i}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\beta_0 + \boldsymbol{\beta' X_i})} e^{\frac{-z^2}{2}} dz, \tag{17}$$

after substitution for the purpose of this study is:

$$F(\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i)} e^{\frac{-z^2}{2}} dz. \tag{18}$$

2 ESTIMATES OF UNKNOWN MODELS PARAMETERS

Estimating the unknown parameters of nonlinear regression models is no problem today. The software uses an iterative method. The software itself selects the initial value. To estimate the parameters of LOGIT and PROBIT model Statgraphics Centurion XVI version 16.1.11 and Gretl 1.8.7 build 2010-01-24 were used.

2.1 LOGIT model, year 1990 versus 1995

Based on the methodology showed above the estimates of unknown parameters of LOGIT models for males in 1990 and 1995 as well as for females in 1990 and 1995 were calculated. Model for males, who were invited to the general medical examination in 1990 is showed in Table 3, model for females, who were invited to the general medical examination in 1990 is showed in Table 4. Using data acquired in 1995, there were calculated estimates of parameters for both males (see Table 5) and females (see Table 6) once again.

Table 3 LOGIT model for males, 1990

Param	Estimate	St. Error	Est. Odds Rat.
C	–15.3537	3.61058	
AGE	0.327231	0.0842361	1.37899
CIRD	0.362252	0.0811236	1.42556
CEL	–8.18332	1.8898887	0.000266545

Source: Own construction

Table 4 LOGIT model for females, 1990

Param	Estimate	St. Error	Est. Odds Rat.
C	–15.7816	3.71599	
AGE	0.329121	0.0853405	1.38975
CIRD	0.361625	0.0857813	1.43566
CEL	–8.19201	1.89216	0.000276858

Source: Own construction

Table 5 LOGIT model for males, 1995

Param	Estimate	St. Error	Est. Odds Rat.
C	–9.17489	1.90414	
AGE	0.162236	0.0322365	1.18122
CIRD	0.189985	0.0488745	1.21113
CEL	–1.85112	0.5899631	0.159863

Source: Own construction

Table 6 LOGIT model for females, 1995

Param	Estimate	St. Error	Est. Odds Rat.
C	–9.38463	1.9487	
AGE	0.161341	0.0391378	1.17509
CIRD	0.195672	0.0471524	1.21613
CEL	–1.84446	0.522231	0.158111

Source: Own construction

We can see that all estimates of unknown parameters are statistically significant. The values from Tables 3–6 can be used to achieve the formulas (3), (10) and (13).

Now we express if the model is estimated well as a whole. As evaluation criteria we can choose for example McFadden's coefficient of determination, as set out e.g. by Freese and Long (2006), whose formula is:

$$McFadden's\ R^2 = 1 - \frac{LLF_{ur}}{LLF_r}, \tag{19}$$

where LLF_{ur} is the value of unlimited maximum likelihood function with all explanatory variables and LLF_r is the value of restricted maximum likelihood function without explanatory variables – a model with constant only. For the purpose of this study can be written as:

$$McFadden's\ R^2 = 1 - \frac{\ln L(b_0, b_1, b_2, b_3,)}{\ln L(b_0)}. \tag{20}$$

We can also use McFadden's corrected coefficient of determination, which also reflects a number of redundant variables in the model. Its formula can be written for the purpose of this study as:

$$McFadden's\ R^2_{adj.} = 1 - \frac{\ln L(b_0, b_1, b_2, b_3,) - K}{\ln L(b_0)}, \tag{21}$$

where K is the number of explanatory variables. McFadden's adjusted coefficient of determination is suitable for comparing of individual models. Table 7 shows the values of these evaluation criteria.

Table 7 Values of model evaluation criteria

LOGIT model for	McFadden's R^2 (in %)	McFadden's $R^2_{adj.}$ (in %)
males 1990	70.2291	65.2763
females 1990	70.3652	65.9664
males 1995	36.2791	31.2406
females 1995	38.3223	33.3214

Source: Own construction

Further evaluation of this model is available in Tables 8–11. In determining whether the model can be simplified, notice that the highest P-value for the likelihood ratio tests is 0.0002. Because the P-value is less than 0.05, it is irrelevant to remove any variable from the model.

Table 8 Likelihood rat. test, males 1990

Factor	Chi-Sq.	Df	P-Value
AGE	32.1326	1	0.0000
CIRD	37.6458	1	0.0000
CEL	77.6632	1	0.0000

Source: Own construction

Table 9 Likelihood rat. test, females 1990

Factor	Chi-Sq.	Df	P-Value
AGE	31.2349	1	0.0000
CIRD	38.8651	1	0.0000
CEL	78.8339	1	0.0000

Source: Own construction

Table 10 Likelihood rat. test, males 1995

Factor	Chi-Sq.	Df	P-Value
AGE	22.6632	1	0.0000
CIRD	23.1131	1	0.0000
CEL	15.5439	1	0.0001

Source: Own construction

Table 11 Likelihood rat. test, females 1995

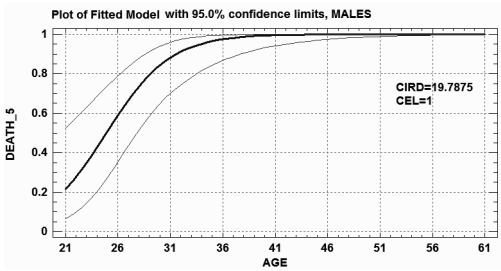
Factor	Chi-Sq.	Df	P-Value
AGE	21.5910	1	0.0000
CIRD	22.1363	1	0.0000
CEL	14.2464	1	0.0002

Source: Own construction

Based on the estimated parameters, there were constructed Figures 2–9, where the vertical axis is the probability of death of x-year old person during next five years after the general medical examination, the horizontal axis expresses a person's age. Value of CIRD in graphs is mean of the population.

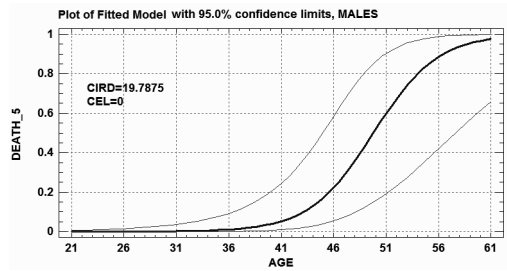
Figure 2 shows the probabilities of death of x-year old males during next 5 years after the general medical examination in 1990, if the males had diagnosed the disease of the digestive system (known as celiac disease now). Figure 3 shows the probabilities of death of x-year old males during next five years after the general medical examination in 1990, if the males had diagnosed any other disease (e.g. diabetes, high blood cholesterol, high blood pressure, bronchial asthma or other).

Figure 2 Probability of death, males, 1990, celiac disease – yes



Source: Own construction

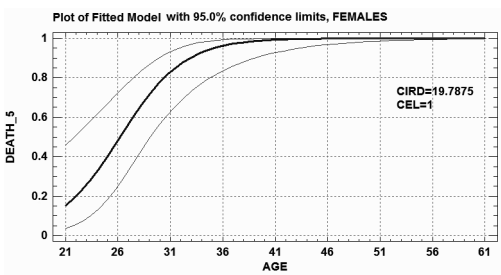
Figure 3 Probability of death, males, 1990, celiac disease – no



Source: Own construction

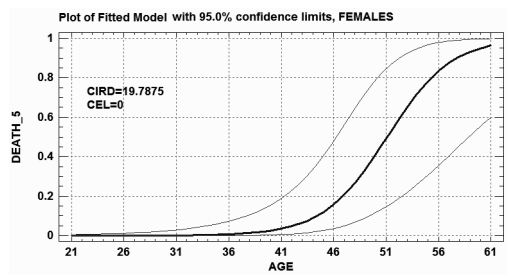
The probabilities for females with disease of the digestive system in 1990 are shown in Figure 4 and the probabilities for females with any other disease (e.g. diabetes, high blood cholesterol, high blood pressure, bronchial asthma or other) are shown in Figure 5.

Figure 4 Probability of death, females, 1990, celiac disease – yes



Source: Own construction

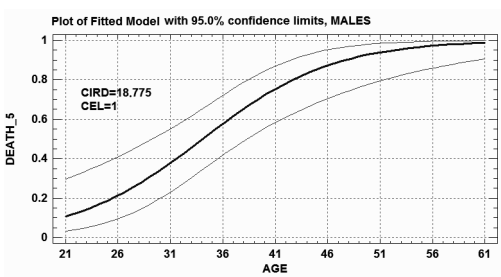
Figure 5 Probability of death, females, 1990, celiac disease – no



Source: Own construction

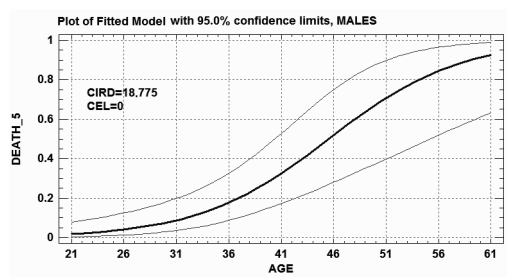
As time went on, the probabilities of death of x-years old person decreased. Decreased for both males and females, but for females a little more. Probability of death of x-year old males during next 5 years after the general medical examination in 1995 (if the males had diagnosed the celiac disease) shows Figure 6. Figure 7 shows the probability of death of x-year old males during next five years after the general medical examination in 1995, if the males had diagnosed any other disease.

Figure 6 Probability of death, males, 1995, celiac disease – yes



Source: Own construction

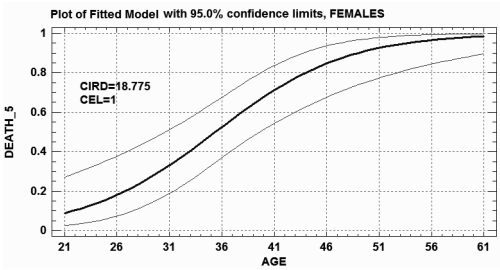
Figure 7 Probability of death, males, 1995, celiac disease – no



Source: Own construction

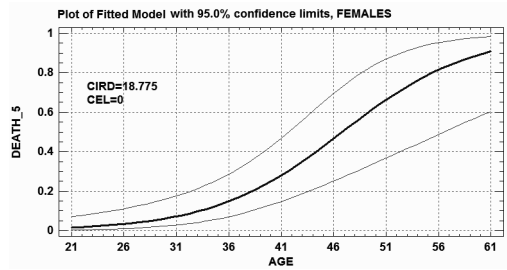
The same situation, but for females, is shown in Figure 8 in case of celiac disease and in Figure 9 in case of any other disease.

Figure 8 Probability of death, females, 1995, celiac disease – yes



Source: Own construction

Figure 9 Probability of death, females, 1995, celiac disease – no



Source: Own construction

2.2 PROBIT model, year 1990 versus 1995

Similarly as in the case of LOGIT model, there were calculated the estimates of the unknown parameters of PROBIT model for the same data. Model for males, who were invited to the general medical examination in 1990 is shown in Table 12, model for females in 1990 is shown in Table 13. Using data acquired in 1995, estimates of parameters for both males (see Table 14) and females (see Table 15) were calculated once again.

Table 12 PROBIT model for males, 1990

Param	Estimate	St. Error
C	–8.7076	1.90321
AGE	0.179963	0.0443266
CIRD	0.212333	0.0441222
CEL	–4.82122	1.0121123

Source: Own construction

Table 13 model for females, 1990

Param	Estimate	St. Error
C	–8.94966	1.95924
AGE	0.1862	0.0458694
CIRD	0.206262	0.0455631
CEL	–4.80235	1.01696

Source: Own construction

Table 14 PROBIT model for males, 1995

Param	Estimate	St. Error
C	–4.85958	0.983271
AGE	0.0900233	0.0236552
CIRD	0.1023366	0.0245888
CEL	–0.987444	0.2796356

Source: Own construction

Table 15 PROBIT model for females, 1995

Param	Estimate	St. Error
C	–4.96564	1.00653
AGE	0.0815844	0.0208831
CIRD	0.11225	0.0252715
CEL	–0.974493	0.282501

Source: Own construction

All estimates of unknown parameters are statistically significant. The values from Tables 12–15 could be used to achieve to the formulas (16) and (18). Table 16 shows the values of models evaluation criteria and from partial likelihood ratio tests is evident (see Tables 17–20), that the models are really estimated as far as possible well.

Table 16 Values of model evaluation criteria

PROBIT model for	McFadden's R^2 (in %)	McFadden's $R^2_{adj.}$ (in %)
males 1990	70.5539	65.6010
females 1990	70.5433	65.1132
males 1995	34.7826	29.7441
females 1995	34.0030	30.3366

Source: Own construction

Table 17 Likelihood rat. test, males 1990

Factor	Chi-Sq.	Df	P-Value
AGE	33.1369	1	0.0000
CIRD	38.6366	1	0.0000
CEL	77.9966	1	0.0000

Source: Own construction

Table 18 Likelihood rat. test, females 1990

Factor	Chi-Sq.	Df	P-Value
AGE	33.0536	1	0.0000
CIRD	39.4311	1	0.0000
CEL	79.5142	1	0.0000

Source: Own construction

Table 19 Likelihood rat. test, males 1995

Factor	Chi-Sq.	Df	P-Value
AGE	18.3636	1	0.0000
CIRD	22.3147	1	0.0000
CEL	13.3289	1	0.0003

Source: Own construction

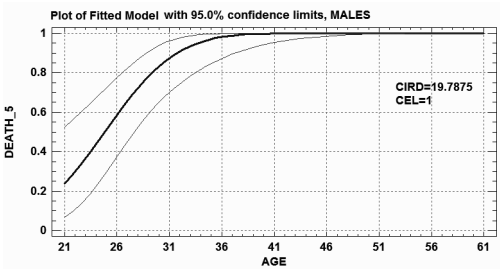
Table 20 Likelihood rat. test, females 1995

Factor	Chi-Sq.	Df	P-Value
AGE	18.6966	1	0.0000
CIRD	21.6042	1	0.0000
CEL	12.4251	1	0.0004

Source: Own construction

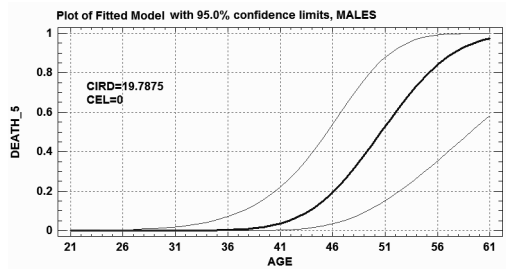
In the same system as in LOGIT models are organized outputs of probabilities of death of x-years old persons. These outputs result from Figures 10–17.

Figure 10 Probability of death, males, 1990, celiac disease – yes



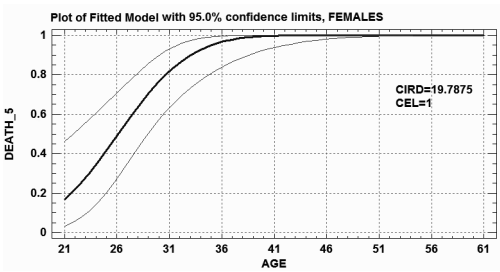
Source: Own construction

Figure 11 Probability of death, males, 1990, celiac disease – no



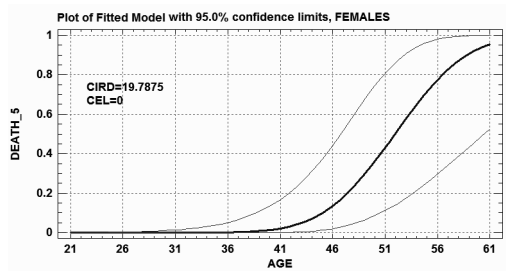
Source: Own construction

Figure 12 Probability of death, females, 1990, celiac disease – yes



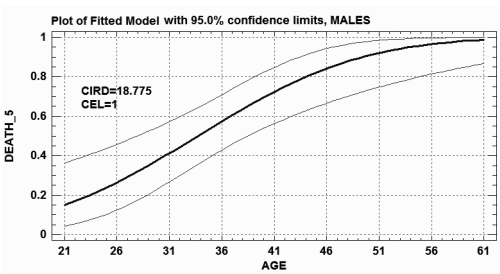
Source: Own construction

Figure 13 Probability of death, females, 1990, celiac disease – no



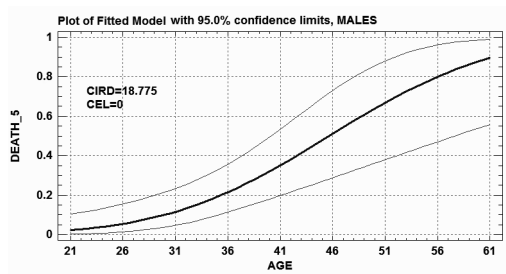
Source: Own construction

Figure 14 Probability of death, males, 1995, celiac disease – yes



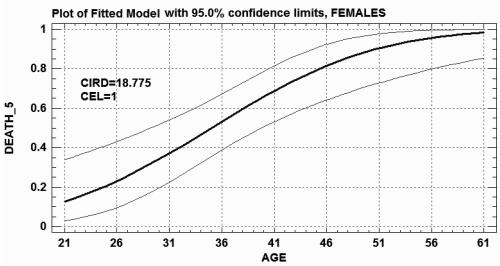
Source: Own construction

Figure 15 Probability of death, males, 1995, celiac disease – no



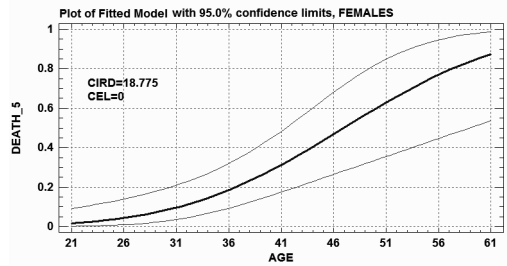
Source: Own construction

Figure 16 Probability of death, females, 1995, celiac disease – yes



Source: Own construction

Figure 17 Probability of death, females, 1995, celiac disease – no



Source: Own construction

CONCLUSION

The aim of this study was to analyse the probability of death of x-year old persons during next five years after the general medical examination in 1990 and 1995. Both analyses were solved using LOGIT and PROBIT models and tried to confirm the hypothesis claiming, that the probability of death of x-year old person suffering from celiac disease decreased few years after the gaining of another medical knowledge from abroad. At the beginning of the independent Czech Republic, the life expectancy at birth of people suffering from celiac disease was quite low, because the detailed diagnosis of this disease from abroad came later. With increasing age, the probability of death of these people grew faster than at the total population. This difference is already much smaller.

LOGIT and PROBIT models in this study do not differ much from each other, the results are almost comparable. Historical data from this last period is not possible to separate and the presented outputs of this study are based on specific research. Even if some assumptions for the application of methods of mathematical statistics are broken, it is possible to say, that the key hypothesis was confirmed.

References

Alergie na lepek (Allergy to Gluten) [online]. Společnost pro bezlepkovou dietu o. s. (Society for the Gluten-Free Diet). <http://celiak.cz/o-nemoci/alergie-na-lepek>.

CZSO. *Naděje dožití při narození v letech 1950–2010* (Life Expectancy at Birth between the Years 1950–2010) [online]. Czech Statistical Office. <http://www.czso.cz/csu/redakce.nsf/i/nadeje_doziti_pri_narozeni_v_letech_1950_2011>.

Diagnostika celiakie (Diagnosis of Celiac Disease) [online]. Poradenské centrum pro celiakii a bezlepkovou dietu (Advice Center for Celiac Disease and Gluten-Free Diet). <http://www.bezlepkovadieta.cz/celiakie-celiakalni-sprue/261-3/diagnostika-celiakie>.

FREESE, J., LONG, J.S. *Regression Models for Categorical Dependent Variables Using Stata*. College Station: Stata Press, 2006.

HOYOS, D., MARIEL, P., MEYERHOFF, J. *Comparing the Performance of Different Approaches to Deal with Attribute Non-Attendance in Discrete Choice Experiments: a Simulation Experiment. BILTOKI 201001,* Universidad del País Vasco – Departamento de Economía Aplicada III (Econometría y Estadística), 2010.

CHRISTENSEN, *R. Log-Linear Models.* New York: Springer-Verlag, 1990.

LOGAN, R. F., RIFKIND, E. A., TURNER, I. D., FERGUSON A. Mortality in Celiac Disease. *Gastroenterology,* 1989, 97(2), pp. 265–271.

PAVLÍK, Z., RYCHTAŘÍKOVÁ, J., ŠUBRTOVÁ, A. *Základy demografie* (Fundamentals of Demography). 1st Ed. Prague: Academia, 1986.

RUBIO-TAPIA, A. et al. Increased Prevalence and Mortality in Undiagnosed Celiac Disease [online]. *Gastroenterology,* 2009, 137, pp. 88–93. [cit. 16.8.2012]. <http://www.celiac.org/images/stories/PDF/Increase%20Prevalence_%20Tapia%20and%20Murray%202009.pdf>.

SPECTOR, L. C., MAZZEO, M. Probit Analysis and Economic Education. *Journal of Economic Education,* Spring, 11, 1980, pp. 37–44.

ŠIMPACH, O. Dopad získání lékařských poznatků na naději dožití u osob s celiakií a srovnání s nadějí dožití celkové populace (The Impact of the Acquisition of Medical Knowledges on the Life Expectancy of People with Celiac Disease and a Comparison with the Life Expectancy of the General Population). Prague, 05.12.2011–06.12.2011. In: *RELIK 2011 Reprodukce lidského kapitálu* (Reproduction of Human Capital). Slaný: Melandrium, 2011, pp. 1–9.

YANG, CH. W., RAEHSLER, R. D. An Economic Analysis on Intermediate Microeconomics: An Ordered PROBIT Model. *Journal for Economic Educators,* Vol. 5, No. 3, 2005.

Human Development and Poverty – a Perspective Across Indian States

Sonu Madan[1] | *I. G. P. G. Regional Centre, Mirpur, M. D. University, Rohtak, Haryana, India*

Abstract

The ultimate objective of development planning and policies is to increase social welfare and well-being of the society. As income alone is an incomplete measure of well-being of any society, human development attempts to capture quantitative as well as qualitative aspects of human well-being by encapsulating indicators of longevity, literacy and a decent standard of living. Human development is about enlarging choices, whereas poverty implies denial to the opportunities and choices most basic to human development. The main concern of this paper is to examine the transformation of development efforts into the well- being of the society, with special reference to India. Here an attempt has been made to find complementary between Human Development Index (HDI) and Multidimensional Poverty Index (MPI) across major Indian states using regression analysis. The negative relationship between the two underlines the need of raising economic and educational opportunities and their equitable distribution among all the sections of the society.

Keywords

Human well-being, human development index, poverty, multi-dimensional poverty index

JEL code

I320, I380

INTRODUCTION

For a long time, development had been conceived as economic development and often related to the level and structure of income. Though economic growth, increasing trade and investment, technological advance - are very important, but development process tends to focus on more and more people rather than mere economic growth of any nation or state in all. Subsequently, focus has shifted to human development (HD), which is about people, about expanding their choices to live full and creative lives with freedom and dignity. The concept of human development has got wide acceptance among academicians, researchers, planners and policy makers and is equally accepted among developing and developed countries. The genesis of the term human development as popularized by the United Nations Development Programme (UNDP) may be found in the writings of the Nobel laureate, Amartya Sen and Mahbub ul Haq. In 1990, the UNDP brought out its first global Human Development Report (HDR). Ever since its publication, under the guidance of Mahbub ul Haq, efforts have been made to devise and further refine the measures of human development (McGillivray, 1991, McGillivray, White 1994, Srinivasan, 1994,

[1] Asst. Professor, I. G. P. G. Regional Centre, Mirpur, M. D. University, Rohtak, Haryana, India. E-mail: sonumadan15@gmail.com, phone: (+91)8685911117.

Streeten, 2000, Neumayer, 2001, Noorbakhsh, 1998a, 1998b, 2002, Malhotra, 2006). The report ranks countries on the basis of composite index, popularly known as Human Development Index (HDI)[2] for most of the countries in the world.

The process of human development can be seen as a 'process of expanding the capabilities of people' (Sen, 1984). Capabilities refer to the alternative combinations of functioning the person can achieve, and from which a person can chose a collection. The notion of freedom is embodied in the term 'capabilities' – the range of options a person has in deciding what kind of a life to lead (Muellbauer, Kanbur, Williams, 1987, Sen, 1992, 1993). Human development has also been defined as the process of 'enlarging the range of people's choices' (UNDP, 1990) as fundamental to expanding human choices is building human capabilities and increasing the range of accessible things. Acquisition of knowledge, the need to lead a long and healthy life and the need to have access to resources required for a decent standard of living have been identified as three essential choices for the attainment of human development. Additional choices valued by people include political, economic and social freedom to opportunities for being creative, productive and also to enjoy personal self-respect and guaranteed human rights. Therefore, the development must be more than just an expansion of income and wealth. Its focus must be people (UNDP, 1990). In this line, the introduction of ethical considerations has been the hallmark of human development approach. Pushing the frontiers of measurement has always been a keystone of this approach. The approach has enabled innovative thinking about progress by capturing the simple but powerful idea that the development is about much more than income. Seen in this perspective, main goal of development is- people free from poverty as high poverty level is synonymous with poor quality of life, deprivation, malnutrition, illiteracy, indicating low level of human development. Over the years the HDI has introduced new measures to evaluate progress in reducing poverty (<http://hdr.undp.org/en/reports/global/hdr2010/summary/measures>).

As far as poverty is concerned, poors' are identified as those unable to get minimum required calorie per day to keep body and soul together. The proportion of population not able to attain the specified level of expenditure is segregated as poor (Bhagawati, 1988). Using such an approach, the Planning Commission, Government of India, has been establishing the head-count ratio (HCR) of the poor at state level, separately for rural and urban areas, for over three decades. The Planning Commission has defined poverty line as a minimum consumption expenditure requirement for an average per capita food energy norm of 2400 and 2100 calories per day for rural and urban areas, respectively (Minhas, Kansal, Jagdish, Joshi, 1986, Bagchi, Choudhury, 1989). The required per capita income, to get minimum food to meet the required calorie, is rupees (Rs.) 328 ($ 8) and Rs.456 ($11) per month in rural and urban areas, respectively, at 1999–2000 prices. That is, a family of five requires at least Rs. 1640 and Rs. 2280 income per month, respectively. The Planning Commission recently (March 20th, 2012) has revised the required per capita income to Rs. 22.42 ($ .44) and Rs. 28.35 ($.56) per day i.e. Rs. 672.6 ($13.32) and Rs. 850.5 ($16.84) per month in rural and urban areas, respectively, to meet minimum consumption expenditure. Those who have less income than this are considered the people living below poverty line (BPL). This is much less than the World Bank's internationally comparable measure of extreme poverty i.e. $ 1.25 / day. The wide difference between national and international measures of income poverty is due to the objective criteria, generally evolved by national policy makers and planners, to measure poverty at a given point of time. Such criteria's differ in context with socio-economic conditions, expenditure required to meet minimum consumption needs, availability of resources and the set objectives to be achieved at a given point of time.

[2] The Human Development Index (HDI) measures the average achievements of a country in basic human capabilities. The HDI indicates whether the people lead a long and healthy life, are educated and knowledgeable and enjoy a decent standard of living. The HDI examines the average condition of all people in a country.

This is the traditional way to capture poverty, which requires two components-distributions of household expenditure and poverty line, and is often linked to the amount of money that households need to fulfil their nutritional requirements. Despite the strong articulation of a multi-dimensional view of human poverty, it is not able to capture the multiple dimensions of poverty.

Poverty prevails everywhere and poverty alleviation has been the main concern of human development. Though human development is much more than mere poverty eradication, still poverty and human development remains two sides of the same coin; one presupposes and challenges the other. Moreover, it is not only poverty, but also the proportion of the poors' and intensity of poverty, which determine success of development programmes. A significant negative relationship between human development and poverty underlines a successful development strategy. Moreover, human development may not have equal effect on all the dimensions of poverty. Hereby, it may help to split out which of dimensions of poverty are significantly affected by the human development and which one has remained untouched. This may further help to revise the methodology to construct HDI.

Objectives of the Study

- To analyse differences in poverty levels across major Indian states.
- To examine the inter-state variation in multiple dimensions of poverty and human development in India.
- To explore the relationship between human development and various dimensions of poverty adopted by global MPI.

Research Hypothesis

H0: Human development and multi-dimensional poverty are independent i.e. Human development is not correlated with any of the parameters of multi-dimensional poverty.

H1: Human development and multi-dimensional poverty are negatively correlated i.e. low level of human development reflects high level of deprivation indicated by the multi-dimensional poverty parameters of health, education and standard of living.

1 REVIEW OF LITERATURE

The real aim of development is not the development of a section, but of whole gamut wherein maximum people are out of vicious trap of poor quality of life, deprivation, malnutrition, illiteracy and low level of human development (HD). Thus, the main goal of development is to free people from poverty as high poverty levels are synonymous with poor quality of life, deprivation, malnutrition, illiteracy indicating low level of human development. Poverty alleviation programmes can not work unless understanding who the poors' are (Mehta, Venkataraman, 2000)? There is, however, consensus at a global level that without eradication of poverty from society human development, in whatever way one defines, remains pretence. This however requires equal distribution of growth benefits among all sections of the society. To this end, the Government of India has been concerned about rising inequalities and uneven distribution of the benefits of growth (Suryanarayana, Agrawal, Prabhu, 2011).

For the successful poverty alleviation programmes, in accordance with development, it is very important to define poverty appropriately. The poverty line, as defined by Planning Commission, considers only minimum nutritional requirements to survive and work, but does not consider other necessities such as fuel, housing, clothing, health services that are also required to survive and work. Moreover, not only the conventional poverty line, but also intensity and dimensions of poverty are a matter of concern. There is a sizeable stratum of the households – casual farm and non-farm labourers, artisans, small and marginal farmers, petty traders, hawkers etc. struggling very hard to live above poverty line. In adverse situations such as illness in family, accident, death or natural calamity like heavy rain, flood, cyclone etc. and / or non-availability of work, their income sink to the level below the poverty line (Anirudh, 2003).

Broadly, poverty status must include certain other dimensions like approachability towards educational, health, housing and sanitation facilities to incorporate a decent standard of living. As the basic purpose of human development is to go beyond mere income resources, the poverty measurement should go far beyond mere inadequate income to encapsulate poor health, inadequate nutrition, low education and skills, bad housing and sanitation conditions, inadequate livelihoods, social exclusion and lack of participation (UNDP, 2010). For this reason, since 1997, Human Development Reports (HDRs) have been measuring poverty in ways different than traditional income based measures. The Human Poverty Index (HPI) was the first such measure which used country averages to reflect aggregate deprivations in health, education and standard of living, was replaced by the Multidimensional Poverty Index (MPI) in 2010 as the former could not identify specific individuals, households or larger groups of people as jointly deprived. The MPI addresses this shortcoming by capturing how many people experience overlapping deprivations and how many deprivations they face on average. The MPI can be broken down by indicator to show how the composition of multidimensional poverty changes for different regions, ethnic groups and so on – with useful implications for policy (Alkire and Santos, 2010a, p.48).

Kakwani and Pernia (2000) defined pro-poor growth as one that enables the poor to actively participate in economic activities and benefit from it significantly. If economic growth, which is an essential component of human development, brings in a sharp increase in inequality, it is possible that the incidence of poverty rises over time because the beneficial effects of growth get offset by the adverse effects of rising inequality, which means that the inequality effect may dominate over the growth effect. Bhagwati (1988) had described this phenomenon as 'immiserizing' growth. Hence, it is important to assess the impact and significance of human development and inequality separately on poverty, which has been attempted in a large number of studies in the past in terms of decomposition exercise (Kakwani, 2000, Jain, Tendulkar, 1990, Sundaram, Tendulkar, 2003).

So far as the relationship between development and poverty is concerned, the study of Kakwani et.al. (2000) brought out the fact that the development process do not benefit the poors' directly, rather the fruits of development received by them remains proportionately less than that of the rich section of the society. Sachs, Bajpai and Ramiah (2002) also observed that the human development across Indian states in the nineties have shown a tendency of divergence rather than convergence, implying that states with a higher per capita income have grown faster than the states with less per capita income. Therefore, it is not only the overall growth but also the composition of growth which is important for poverty reduction.

2 METHODOLOGY

The Planning Commission is a nodal agency for estimating the number and proportion of people living below the poverty lines at national and state level. Hereby, poverty estimations provided by Planning Commission and reports of Working Groups of eminent economists, set up by the Planning Commission, Government of India, have been used to examine the level of poverty in India. Besides, National Sample Survey (NSS) of 61st Round (July 2004 to June 2005) has also been reviewed to access data on consumer expenditure. In order to know various dimensions of poverty, inter-state Multidimensional Poverty Index (MPI) 2010, provided by Oxford Poverty & Human Development Initiative (OPHI), Oxford Department of International Development, Queen Elizabeth House (QEH), University of Oxford, has been used. Besides, UNDP Human Development Index (HDI) 2010 of India and 18 major states have been used to explore the complementary between HDI and MPI. Inter-state HDI is regressed on MPI to obtain the significance of relationship between the two. At the same time HDI is regressed, using step-wise regression method, on various dimensions of MPI i.e. standard of living, education and health dimension of poverty, to explore the significant dimensions affected by HDI.

3 RESULTS AND DISCUSSION

3.1 Poverty Estimates in India

India has been overwhelmingly concerned with income poverty since early in the 1960s when a Working Group of eminent economists was set up by the Planning Commission, Government of India, to assess the extent of poverty in the country. This Group used a nationally desirable minimum level of consumption expenditure to define India's poverty line and based it on a standard balanced diet prescribed by the Nutrition Advisory. The estimation of conventional income poverty was revamped based on the recommendations of the 'Report of the Task Force on Projection of Minimum Needs and Effective Consumption Demand' 1979 and later modified on the basis of the recommendations of the 'Report of the Expert Group on Estimation of Proportion and Number of Poor' 1993. Subsequent studies on poverty in India continued to use either income or consumption as a basis for defining and measuring poverty (Minhas, 1970, Bardhan, 1970, Ahluwalia, 1978).

Currently, key features of poverty estimates in India are the following:

- The poverty measure is a head-count ratio (HCR) based on expenditure poverty line.
- Poverty line is based on the absolute measure of poverty, not on relative measure.
- The starting point for estimating the poverty line is a normative nutritional requirement per person a day at some base point i.e. 2 400 calories per person a day in rural areas and 2 100 calories per person a day in urban areas.
- Nutritional requirements are translated into monetary terms to arrive at a certain level of household / per capita expenditure to obtain the desired calories serves as the poverty line.
- The poverty line over the time period is adjusted keeping in consideration the price variation and selecting an appropriate price deflator.

On the basis of National Sample Survey (NSS) data on consumer expenditure of 61st Round (July 2004 to June 2005), the poverty ratio at the national level is estimated as 28.3 per cent in the rural areas, 25.7 per cent in the urban areas and 27.5 per cent for the country as a whole in 2004–05 using the Uniform Recall Period (URP).[3] The poverty estimates in 2004–05 i.e. 27.5 per cent is comparable with the poverty estimates of 1993–94, which was 36 per cent (Planning Commission, Annual Report, 2010–11). The Planning Commission used to upgrade the per capita expenditure periodically, required to meet the above stated consumption requirements, adopting specific consumer price indices. In this line, two committees are constituted one under the chairmanship of Prof. Suresh D. Tendulkar to conduct BPL census in rural areas and the other under the chairmanship of Prof. S. R. Hasim for the identification of BPL families in urban areas (Kapila, 2011). Hereby, as per Expert Group Report, submitted in December 2009, all India rural poverty head-count ratio for 2004–05 was estimated at 41.8 per cent, 25.7 per cent in urban and 37.2 per cent in rural areas. The report estimated Rs. 446.68 and Rs. 578.8 as poverty line in rural and urban areas respectively (Table 1).

Further, as per Tendulkar Committee recommendations, the state wise urban poverty lines of 2004–05 are updated for 2009–10 based on price rise during this period using Fisher Price Indices.[4] The state wise rural-urban price differential in 2009–10 has been applied on state specific urban poverty lines to get state specific rural poverty lines. As per the revised figures, the all-India HCR has declined by 7.4 per-

[3] Uniform Recall Period (URP): Under URP based poverty estimation methodology, consumer expenditure data for all the items are collected from 30-day recall period.

[4] Fisher Price Indexes- for states relative to All-India rural prices, for states relative to All-India urban prices and within state, rural relative to urban prices, are calculated for 15 commodity groups (as in the NSS Consumption Expenditure Survey): cereals, pulses, milk, oil, egg-fish-meat, vegetables, fresh-fruit, dry-fruit, sugar, salt-spices, other-food, intoxicants, fuel-light, clothing, footwear. At this stage all population figures used for states and sectors are census populations as given by Planning Commission (Planning Commission, 2009, p. 20).

Table 1 Poverty Estimates in India

Poverty Estimates / Year	1993–94			1993–94			2009–10		
	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total
Planning Commission* (URP based Estimates)	37.3	32.4	36.0	28.3 (21.8)	25.7 (21.7)	27.5 (21.8)	-	-	-
Expert Group Report** (MRP based estimates)[5]	50.1	31.8	45.3	41.8	25.7	37.2	33.8	20.9	29.8

Note: Figures in Parenthesis Shows Mixed Recall Period (MRP) Based Poverty Estimates.
Source: *Planning Commission (2011, pp.11–12), **Planning Commission (2009, p. 17), **Planning Commission (2012, p. 6)

centage points (from 37.2 per cent in 2004–05 to 29.8 per cent in 2009–10), with rural poverty declining by 8.0 percentage points (from 41.8 per cent to 33.8 per cent) and urban poverty declining by 4.8 per cent points (from 25.7 per cent to 20.9 per cent).

The figures on poverty, showing a decline in it, may be misleading until and unless the poors (people living below poverty live) from all social strata have not been equally benefited by the downward trend. Among poors, the worse sufferers may continue to be from the socially disadvantaged groups e.g. scheduled castes, scheduled tribes, women and their process of coming out of the poverty line might be slower than other groups. Here, it is worth to correlate poverty and overall human development of people. Moreover, these poverty estimates provides, per-capita income/consumption expenditure based, head-count ratio and do not provide any information on various dimensions of poverty.

3.2 Human Development in India

The Human Development Index, 2010 of India and major states, is shown in Table 2. India has a HDI value of 0.504 (Suryanarayana, Agrawal and Prabhu, 2011, p.8). The highest HDI is recorded for Kerala (0.625) followed by Punjab (0.569) and the lowest for Orissa (0.442), preceded by Bihar (0.447) and Chhattisgarh (0.449). The overall HDI score across Indian states shows a variation of 0.183 ranging between 0.442 (Orissa) and 0.625 (Kerala). Among dimensions of HDI, education and health dimensions have shown a greater degree of variation than that of income dimension (Table 2).

In terms of education dimension of HDI, 11 states have not been able to attain an average score (0.400) and needs special concern towards raising educational facilities by their respective states. In this concern, Kerala again ranks first (0.534) and Rajasthan stands at the last (0.333) showing a variation of 0.201 points. Inter-state variation in health dimension of HDI is maximum i.e. 0.201 points, with Kerala (0.854) at the top position and Madhya Pradesh and Chhattisgarh (0.601) at the last among major states. Income dimension of HDI reveals that 9 out of the 18 major states have better income index than the nation as a whole (0.465). Kerala comes up with first rank (0.535) and Bihar (0.398) proceeded by Orissa (0.400) stands at the last. Inter-state variation in income index is observed to be minimum i.e. 0.137. The point to be noticed is that the position of India in health dimension of HDI is better than that of income and education dimension because of favourable environmental conditions.

---

[5] Mixed Recall Period (MRP): In MRP based poverty estimates, consumer expenditure data for five non-food items, namely clothing, footwear, durable goods, education and institutional medical expenses, are collected for a 365-day recall period and the consumption data for the remaining items are collected for a 30-day recall period.

[6] HDI: The HDI is geometric mean of the Dimension Indices of income, education and health (using the goalposts discussed in HDR 2010 discussed in Suryanarayana, Agrawal, Prabhu, 2011, pp. 26–27).
$HDI = \sqrt[3]{income\ index \times education\ index \times healt\ index}$

Globally, India ranks 119 out of 169 countries with a global HDI value of 0.504 and falls in the category of countries with Medium HD.[7] It falls short of the world average, which is 0.624 (UNDP, 2010, p. 155). The Indian states fall either in the category of Medium HD or Low HD as per the HDR 2010 classification. Kerala, with a global HDI of 0.625, is in the 'Medium HD' category. Other major states in this group are Punjab, Himachal Pradesh, Haryana, Maharashtra, Tamil Nadu, Karnataka, Gujarat and West Bengal. Other nine states, namely Andhra Pradesh, Assam, Uttar Pradesh, Rajasthan, Jharkhand, Madhya Pradesh, Chhattisgarh, Bihar and Orissa fall in the 'Low HD' category.

Table 2 Human Development Index, 2010: India and States

India / States	Dimensions of HDI			HDI	Rank
	Income Index	Education Index	Health Index		
Kerala	0.535	0.534	0.854	0.625	1
Punjab	0.523	0.452	0.782	0.569	2
Himachal Pradesh	0.499	0.468	0.744	0.558	3
Maharashtra	0.489	0.453	0.747	0.549	4
Haryana	0.513	0.432	0.731	0.545	5
Tamil Nadu	0.486	0.454	0.731	0.544	6
Gujarat	0.484	0.403	0.698	0.514	7
West Bengal	0.468	0.397	0.710	0.509	8
Karnataka	0.461	0.396	0.717	0.508	9
Andhra Pradesh	0.467	0.347	0.703	0.485	10
Assam	0.442	0.392	0.616	0.474	11
Rajasthan	0.462	0.333	0.665	0.468	12
Uttar Pradesh	0.444	0.365	0.633	0.468	12
Jharkhand	0.421	0.361	0.658	0.464	14
Madhya Pradesh	0.431	0.355	0.601	0.451	15
Chhattisgarh	0.420	0.358	0.601	0.449	16
Bihar	0.398	0.340	0.658	0.447	17
Orissa	0.400	0.345	0.627	0.442	18
India	**0.465**	**0.400**	**0.688**	**0.504**	

Source: Suryanarayana, Agrawal, Prabhu (2011, pp. 17–18)

Human development presupposes smaller proportion of the 'poors' among its achievements. Hereby, the states with better HDI (categorised as Medium HD) should have lower percentage of people living below poverty line than the states with Low HD. Thus, human development needs to be analysed in terms of multidimensional poverty indicators. While the HDI measures the achievement in the average progress, the MPI measures deprivation in human development as a focus on deprivation is fundamental to human development.

3.3 Multidimensional Poverty Index – A New Perspective to Measure Poverty

The Human Development Report, 2010 presented Multidimensional Poverty Index (MPI), covering 104 developing countries. The MPI presented in the 2011 adds five additional countries (Alkire and Santos, 2011, p. 4) using the same dimensions, indicators, cut-offs, and weights in as in 2010 and updated MPI values for all countries as per newly released data. The MPI is an international measure of acute poverty which identifies deprivations across health, education and standard of living that people face at the same

[7] Countries with HDI value in the range 0.788–0.938 are categorised as countries with Very High HD; Countries with HDI value in the range 0.677–0.784 are categorised as countries with High HD; Countries with HDI value in the range 0.488–0.669 are categorised as countries with Medium HD; Countries with HDI value in the range 0.140–0.470 are categorised as countries with Low HD (Suryanarayana, Agrawal, Prabhu, 2011, pp. 22–23).

time at the household level, and shows the number of multi-dimensionally poor people. (<http://www.ophi.org.uk/multidimensional-poverty-index/mpi-faqs>).

The MPI combines two sets of data to measure poverty (Alkire and Santos, 2011):
First one is the *prevalence* of poverty, or the proportion of people (within a given population) who experience multiple deprivations, also known as *multidimensional head-count ratio* (*H*):

$$H = q / n , \tag{1}$$

here q is the number of people who are multi-dimensionally poor and n is the total population. *The second* component is the *intensity* of their deprivation – the average proportion of (weighted) deprivations they experience:

$$A = \sum_{i=1}^{n} Ci\,(k) / q , \tag{2}$$

where $Ci(k)$ is the censored deprivation score of individual i and q is the number of people who are multi-dimensionally poor. The MPI is the product of both: MPI = $H \times A$.

The Multidimensional Poverty Index (MPI), presented by Human Development Report 2011, covering 109 countries, complements income based poverty measures. It has three dimensions mirroring the HDI – health,[8] education[9] and standard of living[10] (Alkire, Santos, 2011, p. 5) – which are reflected in 10 indicators of multi-dimensional poverty and intensity of deprivations at the household level, each with equal weights within its dimension. It has been estimated to reflect the deprivational perspective of development. Deprivation in health is captured essentially through the nutritional level and child mortality. Deprivation in educational attainments is captured through years of schooling and children enrolled. Similarly, to capture a decent standard of living, six indicators namely cooking fuel, toilet, water, electricity, floor and assets are considered – a household is multi-dimensionally poor if it is deprived in at least two to six indicators, depends on the weight of the specific indicator in the overall measure making 1/3 of the total weight. In other words, a person is identified as poor if he or she has a deprivation score higher than or equal to 1/3 of the (weighted) considered indicators. (UNDP 2010, p. 95). Eight of the ten indicators are connected to Millennium Development Goals (MDG)[11] indicators.

[8] Health: Health dimension includes two parameters weighted equally at 1/6. A) Child Mortality: Deprived if any child has died in the family. B) Nutrition: deprived if any adult or child from whom there is nutritional information is malnourished.

[9] Education: Education dimension includes two parameters weighted equally at 1/6. A) Years of Schooling: Deprived if no household member has completed five years of schooling. B) School Attendance: Deprived if any school age child is not attending school in years 1–8.

[10] Standard of Living: Standard of living dimension includes six parameters weighted equally at 1/18. A) Electricity: Deprived if household has no electricity. B) Drinking Water: Deprived if household has no access to clean drinking water or clean water is 30 minutes' walk from home. C) Sanitation: Deprived if household not having access to adequate sanitation. D) Flooring: Deprived if household has a dirt, sand or dung floor. E) Cooking Fuel: Deprived if household is using dung, charcoal or wood as cooking fuel. F) Assets ownership: Deprived if household does not more than one of radio, TV, telephone, bicycle, motorcycle, refrigerator, and does not own a car or a tractor.

[11] MDGs are a set of numerical and time-bound targets to measure achievements in human and social development. A) Goal-1: Eradicate extreme poverty and hunger. B) Goal-2: Achieve universal primary education. C) Goal-3: Promote gender equality and empower women. D) Goal-4: Reduce child mortality. E) Goal-5: Improve maternal health. F) Goal-6: Combat HIV/AIDS, Malaria and other diseases. G) Goal-7: Ensure environmental sustainability. H) Goal-8: Develop a global partnership for development.

The other two indicators[12] (flooring and electricity) provide some rudimentary indication of the quality of housing.

As far India is concerned, poverty estimates provided by Oxford Poverty and Human Development Initiative (OPHI) of Oxford University and the Human Development Report Office of the United Nations Development Programme (UNDP) in July, 2010 provides an insight that 55% of the Indian population is poor – deprived in 30% indicators. It is much higher than the official figure of 32.7%. About 39% population is poor in 40% indicators; 30% Indians are poor in 50% indicators, 20% people are deprived on 60% indicators, and 10% population is deprived on 70% of the 10 indicators.

The same results are presented in Table 3 showing the multi-dimensional poverty reflected through various indicators (Table 3). It is clear that about 52% population is deprived of cooking fuel, 49% people lack proper sanitation, and 39% are undernourished. The MPI analysis also reveals that three largest deprivations in India are: Nutrition (biggest) followed by school enrollment and child mortality.

The figures on MPI shown in Table 4 give a clear picture of severity of the problem of poverty, especially in Bihar, Jharkhand, Madhya Pradesh, Chhattisgarh, Uttar Pradesh, Rajasthan, Orissa and West Bengal. These eight states occupy last ranks, assigned to the states in ascending order of multidimensional poverty, i.e. 11–18 (Table 4). The 'MPI poor' people in these states are even more than that of in the 26 poorest African countries combined (410 million). The above stated states have also shown their relatively poor performance on human ground also as indicated by their positioning in terms of HDI.

Table 3 MPI in India: Dimensions and Indicators

Dimensions of MPI	Indicators of MPI	Occurrence of poverty indicators (in %)	Contribution of Indicators to the MPI (in %)	Relative weight	Connection with to MDG indicators
Education	Schooling	18	10	1/6 = 16.7%	MDG 2
	Enrolment	25	14	1/6 = 16.7%	MDG 2
Health	Child Mortality	23	13	1/6 = 16.7%	MDG 4
	Nutrition	39	22	1/6 = 16.7%	MDG 1
Standard of Living	Electricity	29	5	1/18 = 5.6%	MDG 7
	Sanitation	49	9	1/18 = 5.6%	MDG 7
	Drinking Water	12	2	1/18 = 5.6%	MDG 7
	Floor	40	8	1/18 = 5.6%	MDG 7
	Cooking Fuel	52	10	1/18 = 5.6%	MDG 7
	Assets	38	7	1/18 = 5.6%	MDG 7

Source: Alkire, Santos (2010a, p. 17), Alkire, Santos (2010b, pp. 3–4)

3.4 Interconnection between Human Development and Multidimensional Poverty

In an attempt to find statistical relationship between HDI and MPI, Table 5 summarises the results and confirms a significant negative relationship between the two. The negative value of coefficient of correlation i.e. –.899 confirms the negative relationship between HDI and MPI and is significant at 0.01 levels (Table 5). The value of t and F statistics also confirms the significance of MPI in explaining HDI. High poverty levels have resulted in low level of human development.

[12] Flooring and electricity are the only indicators not explicitly listed as MDG indicators. However, they are closely related to MDG 7 – Achieve Environmental Sustainability. As access to safe drinking water serves directly to satisfy the need of hydration and hygiene, hereby hygiene is also facilitated by the access flooring material. Electricity being a safer means of lighting, allows people to be independent during the night time, contributes to a healthy home environment.

Table 4 MPI – Components, Dimensions and Indicators and HDI – India and States, 2011

India / States	MPI	MPI Rank	Components of MPI		Education Dimension of MPI		Percentage Deprivations in Education	Health Dimension of MPI		Percentage Deprivations in Health	Standard of Living Dimension of MPI						Percentage Deprivations in Standard of Living	HDI*	Rank
			H (Proportion of poor)	A (Average intensity of deprivations)	Schooling	School Attendance		Mortality	Nutrition		Electricity	Sanitation	Drinking Water	Floor	Cooking Fuel	Assets			
Kerala	**0.065**	1	0.159	0.409	0.01	0.07	20.3	0.04	0.12	40.4	0.05	0.04	0.09	0.03	0.15	0.11	39.3	0.625	1
Punjab	0.120	2	0.262	0.460	0.08	0.13	30.0	0.09	0.17	36.1	0.02	0.20	0.01	0.16	0.23	0.11	33.9	0.569	2
Himachal Pradesh	0.131	3	0.310	0.423	0.04	0.07	13.6	0.09	0.25	43.3	0.01	0.28	0.08	0.15	0.19	0.20	43.1	0.558	3
Tamil Nadu	0.141	4	0.324	0.436	0.09	0.08	19.4	0.11	0.21	37.5	0.07	0.31	0.05	0.12	0.30	0.24	43.2	0.544	6
Maharashtra	0.193	5	0.401	0.481	0.08	0.15	20.0	0.14	0.30	37.8	0.13	0.36	0.08	0.27	0.34	0.28	42.2	0.549	4
Haryana	0.199	6	0.416	0.479	0.08	0.20	23.8	0.15	0.30	37.6	0.08	0.34	0.08	0.24	0.39	0.25	38.6	0.545	5
Gujarat	0.205	7	0.415	0.492	0.12	0.13	20.3	0.17	0.33	40.6	0.09	0.36	0.10	0.24	0.36	0.29	39.2	0.514	7
Andhra Pradesh	0.211	8	0.447	0.471	0.19	0.13	25.1	0.16	0.29	35.0	0.08	0.41	0.06	0.19	0.42	0.35	39.9	0.485	10
Karnataka	0.223	9	0.461	0.483	0.12	0.21	24.9	0.17	0.33	36.9	0.08	0.41	0.12	0.19	0.42	0.32	38.2	0.508	9
Assam	0.303	10	0.576	0.525	0.19	0.21	22.0	0.19	0.37	31.0	0.41	0.45	0.23	0.50	0.55	0.42	47.0	0.474	11
West Bengal	0.317	11	0.583	0.543	0.25	0.23	25.4	0.19	0.42	32.0	0.39	0.47	0.07	0.48	0.57	0.43	42.7	0.509	8
Orissa	0.345	12	0.640	0.540	0.23	0.19	20.3	0.24	0.45	33.3	0.43	0.62	0.20	0.51	0.63	0.39	46.4	0.442	18
Rajasthan	0.351	13	0.642	0.547	0.21	0.32	25.0	0.28	0.44	34.2	0.31	0.60	0.24	0.36	0.61	0.47	40.8	0.468	12
Uttar Pradesh	0.386	14	0.699	0.552	0.18	0.36	23.4	0.37	0.46	35.8	0.48	0.62	0.07	0.58	0.66	0.41	40.8	0.468	12
Chhattisgarh	0.387	15	0.719	0.539	0.21	0.29	21.6	0.31	0.52	35.8	0.24	0.69	0.22	0.64	0.70	0.48	42.5	0.449	16
Madhya Pradesh	0.389	16	0.695	0.560	0.22	0.32	22.9	0.31	0.50	34.7	0.25	0.65	0.31	0.57	0.67	0.52	42.4	0.451	15
Jharkhand	0.463	17	0.770	0.602	0.26	0.45	25.3	0.30	0.56	31.0	0.55	0.73	0.42	0.63	0.76	0.55	43.7	0.464	14
Bihar	0.499	18	0.814	0.613	0.35	.52	29.0	0.35	0.61	32.0	0.65	0.74	0.04	0.70	0.79	0.57	39.0	0.447	17
India	**0.296**		**0.554**	**0.535**	**0.18**	**0.25**	**24.0**	**0.23**	**0.39**	**34.7**	**0.29**	**0.49**	**0.12**	**0.40**	**0.52**	**0.38**	**41.3**	**0.504**	

Source: Alkire, Santos (2010a, pp. 124–125)

Table 5 HDI and MPI – Model Statistics

Model		Unstandardized Coefficients	Std. Error	Standardized Coefficients	t	R	R Square	F	Durbin-Watson
		B		Beta					
1	(Constant)	.605	.014		44.689**				
	MPI	–.371	.045	–.899	–8.210**	–.899*	.808	67.407**	1.385

Note: A) Predictors: (Constant), MPI, B) Dependent variable: HDI, *Significant at 0.01 levels (2 tailed), **Significant at 0.00 level.
Source: Obtained from the information provided in Table 4 using SPSS 12.0

In order to understand the relative significance of various dimensions of poverty in determining human development, step-wise regression has been used to develop three models (Table 6). The model-1, estimates the effect of standard of living dimension of MPI, and then model-2 and model-3 adds education and health dimension of MPI respectively. The information on change in R Square explains the effectiveness of the variables added in the subsequent models. Variable entered in model-1 i.e. standard of living dimension of MPI, predicts only 18 per cent variation in HDI, whereas model-2 with added education dimension of MPI, predicts 52 per cent variation. Variable added in model-2 has significantly changed the values of R Square and F and is significant at .01 levels of significance. It underlines the fact that the standard of living and education dimension of MPI are helpful in determining the value of HDI up to 52 per cent. Model-3 explains the relative significance of health dimension of MDI, which is not found statistically significant as this dimension explains only .047 per cent variation in determining HDI value.

Table 6 HDI and Parameters of Multi-dimensional Poverty – Model Summary

Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change	Durbin-Watson
1	.424	.180	.0483	.180	3.503	.080	
2	.723	.523	.0380	.343	10.780	.005	
3	.755	.569	.0374	.047	1.515	.239	1.143

Note: A) Predictors: (Constant), Standard of living dimension of MPI, B) Predictors: (Constant), Standard of living dimension of MPI, Education dimension of MPI, C) Predictors: (Constant), Standard of living dimension of MPI, Education dimension of MPI, Health dimension of MPI, D) Dependent variable: HDI.
Source: Obtained from the information provided in Table 4 using SPSS 12.0

Table 7 provides the statistical significance of variables excluded from each variable. Model-2 provides information on the excluded variable i.e. health dimension of MPI, whose t value 1.231 is not statistically significant at .05 level of significance and hereby does not meet the criteria for inclusion, so estimation stops at model-2 with two variables as predictors (Table 8).

Table 7 Excluded Variables

Model		Beta In	t	Sig.
1a	Education dimension of MPI	–.689	–3.283	.005
	Health dimension of MPI	.623	3.313	.005
2b	Health dimension of MPI	15.672	1.231	.239

Note: A) Predictors in the model: (Constant), Standard of living dimension of MPI, B) Predictors in the model: (Constant), Standard of living dimension of MPI, Education dimension of MPI. Dependent variable: HDI.
Source: Obtained from the information provided in Table 4 using SPSS 12.0

Table 8 HDI and Coefficients Associated with the Dimensions of MPI

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Correlations (Zero-order)	VIF
		B	Std. Error	Beta				
1	(Constant)	.795	.156		5.094	.000		
	Standard of living dimension of MPI	–7.060	.004	–.424	–1.872	.080	–.424*	1.000
2	(Constant)	1.261	.188		6.716	.000		
	Standard of living dimension of MPI	–1.312	.003	–.788	–3.750	.002	–.424*	1.386
	Education dimension of MPI	–9.428	.003	–.689	–3.283	.005	–.274	1.386

Note: A) Dependent variable: HDI, *Significant at 0.05 level.
Source: Obtained from the information provided in Table 4 using SPSS 12.0

The value of coefficients shows the relative effectiveness of each dimension of MPI in determining HDI (Table 8). Standard of living dimension of MPI is significantly negatively correlated with HDI. The value of t-statistics associated with this is not significant at 0.05 levels as it does not explain, as a single variable, significant variation in HDI (model-1). The same variable when associated with education dimension of MPI explains its significant negative effect on HDI. Both dimensions included in model-2 i.e. standard of living and education dimensions of MPI explain the negative impact of poverty in determining the value of HDI.

CONCLUSION

Poverty head-count ratio in India, a conventional measurement of poverty has reduced by 8.5 percentage points, from 36.0 per cent 1993–94 to 27.5 per cent in 2004–05, as per figures provided by the Planning Commission. The revised figures in this regard, provided by Expert Group, Planning Commission, though present different figures but confirm a decline in this. The poverty head-count ratio, which was 45.3 per cent in 1993–94 reduced to 37.2 percent in 2004 and further estimated 29.8 per cent in 2009–10. These data has provided clear-cut evidence on a decline in poverty in India and seems to correspond with development. The same figure when explored with a new perspective, which is multi-dimensional poverty, brings out anti-development growth process of the economy and calls for immediate attention of policy planners. The MPI represents significant progress in the measurement of poverty in an internationally comparable way, wherein attention has shifted from solely income to include other essentially important dimensions. The inclusion of some essential indicators under various dimensions provides an insight that 55 per cent of the Indian population is poor. These many poor people are a treat to the harmony of any society as the ultimate objective of development planning is human development or increased social welfare and well-being of the people. In an attempt to find the role of poverty as a determinant of HD, regression analysis confirms the negative relationship between the two. Hereby, we may reject null hypothesis as the analysis brings out the fact that among poverty dimensions, standard of living and education dimensions are found to be the significant determinants of HD. At the same time the study indicates that the fruits of development have not been distributed equally among all the persons in society as even the states with better HDI have not shown much advantage over the states with high poverty levels as indicated by their MPI. It underlines the need of expansion of economic activities and educational facilities along with their equitable distribution. Thus, it can be concluded that to raise the level of human development concrete efforts at grass root level, not only towards raising economic resources and developmental opportunities but also to ensure their equal distribution, should be made.

References

AHLUWALIA, M. S. Rural Poverty and Agricultural Performance in India. *Journal of Development Studies*, 1978, 14(13), pp. 298–323.

ALKIRE, S., SANTOS, M. E. *Acute Multidimensional Poverty: A New Index for Developing Countries.* Oxford Poverty & Human Development Initiative (OPHI), Working Paper No. 38, 2010a.

ALKIRE, S., SANTOS, M. E. *India Country Briefing* [online]. Oxford Poverty & Human Development Initiative (OPHI) Multidimensional Poverty Index Country Briefing Series, 2010b. [cit. 20.8.2012]. <www.ophi.org.uk/policy/multidimensional-poverty-index/mpi-country-briefings>.

ALKIRE, S., SANTOS, M. E. *The Multidimensional Poverty Index, Training Material For Producing National Human Development Reports* [online]. Oxford Poverty & Human Development Initiative (OPHI), 2011. [cit. 20.8.2012]. <http://www.ophi.org.uk/wp-content/uploads/OPHI-RP-31a.pdf.pdf?cda6c1>.

ANIRUDH, K. Falling into Poverty: Other Side of Poverty Reduction. *Economic and Political Weekly*, 2003, 38(6), pp. 533–542.

BAGCHI, A., CHOUDHURY U. D. R. Poverty Measures as an Index of Backwardness and their Relevance to Tax Devolution. *Economic and Political Weekly*, 1989, 14(23), pp. 1831–36.

BARDHAN, P. K. On the Minimum Level of Living and the Rural Poor. *Indian Economic Review, New Series*, 1970, 5(1), pp. 129–136.

BHAGAWATI, J. N. Poverty and Public Policy. *World Development*, 1988, 16(5), pp. 539–654.

JAIN, L. R., TENDULKAR, S. D. Role of Growth and Distribution in the Observed Change in Headcount Ratio Measure of Poverty: A Decomposition Exercise for India. *Indian Economic Review*, 1990, 25(2), pp. 165–205.

KAKWANI, N., PERNIA, E. M. What is Pro-Poor Growth? *Asian Development Review*, 2000, 18(1), pp. 1–16.

KAKWANI, N., BRAHM PRAKASH, HYUN SON. Growth, Inequality, and Poverty: An Introduction. *Asian Development Review*, 2000, 18(2), pp. 1–21.

KAKWANI, N. On Measuring Growth and Inequality Components of Poverty with Application to Thailand. *Journal of Quantitative Economics*, 2000, 16(1), pp. 67–79.

KAPILA, U. *Indian Economy-Performance and Policies.* Academic Foundation, New Delhi, 2011, pp. 180–184.

MALHOTRA, R. Towards Implementing the Right to Development: A Framework for Indicators and Monitoring Methods. In ANDREASSEN, A. B., MARKS, S. P. Eds. *Development as a Human Right*, Harvard University Press, Cambridge, 2006, pp. 203–26.

MCGILLIVRAY, M. The Human Development Index-Yet Another Redundant Composite Development Indicator? *World Development,* Elsevier, 1991, No. 19, pp. 1461–68.

MCGILLIVRAY, M., WHITE, H. Measuring Development? The UNDP's Human Development Index. *Journal of International Development,* John Wiley and Sons, 1994, 5(2), pp. 183–92.

MEHTA, J., VENKATARAMAN, S. Poverty Statistics. *Economic and Political Weekly,* 2000, 35(27), pp. 2377–85.

MINHAS, B. S., KANSAL, S. M., JAGDISH, K., JOSHI, P. D. On Reliability of the Available Estimates of Private Consumption Expenditure in India. *The Journal of Income and Wealth,* 1986, 9(2).

MINHAS B. S. Rural Poverty, Land Redistribution and Development Strategy. *Indian Economic Review,* 1970, 5(1), pp. 97–128.

MUELLBAUER, J., KANBUR, R., HART, K., WILLIAMS, B. The Standard of Living: Lecture II, Lives and Capabilities. In SEN, A., Eds. *The Standard of Living.* The Tanner Lectures, 1985, Cambridge University Press, Cambridge, 1987, pp. 20–38.

NEUMAYER, E. The Human Development Index and Sustainability – A Constructive Proposal. *Ecological Economics,* Elsevier, 2001, 39(2), pp. 101–14.

NOORBAKHSH, F. Human Development and Regional Disparities in Iran: A Policy Model. *Journal of International Development,* John Wiley and Sons, 2002, 14(2), pp. 927–49.

NOORBAKHSH, F. A Human Development Index: Some Technical Issues and Alternative Indices. *Journal of International Development,* John Wiley and Sons, 1998b, 10(5), pp. 589–605.

NOORBAKHSH, F. A Modified Human Development Index. World Development, Elsevier, 1998a, 26(3), pp. 517–32.

PLANNING COMMISSION. *Economy and the Plan: An Overview.* Annual Report, Chapter-2, New Delhi: Government of India, 2011.

PLANNING COMMISSION. *Press Note on Poverty Estimates,* New Delhi: Government of India, 2012.

PLANNING COMMISSION. *Report of the Expert Group to Review the Methodology for Estimation of Poverty* [online]. New Delhi: Government of India, 2009. [cit. 25.7.2012]. <http://planningcommission.nic.in/reports/genrep/rep_pov.pdf>.

SACHS, J. D, BAJPAI, N., RAMIAH, A. Understanding Regional Economic Growth in India. *CID Working Paper,* No. 88, Center for International Development, Harvard University, 2002.

SAMPLE REGISTRATION SYSTEM (SRS). Office of Registrar General and Census Commissioner. *Planning Commission and Economic Survey,* New Delhi, 1999–2000.

SEN, A. Capability and Well-Being. In NUSSBAU, M., SEN, A. *The Quality of Life.* A Study Prepared for the World Institute for Development Economics Research (WIDER) of the United Nations University, New Delhi: Oxford University Press, 1993, pp. 30–53.

SEN, A. *Inequality Reexamined, Russell Sage Foundation.* Oxford, New York: Harvard University Press, 1992, pp. 102–114.

SEN, A. *Resources, Values & Development.* New Delhi: Oxford University Press, 1984, p. 497.

SRINIVASAN, T. N. Human Development: A Paradigm or Reinvention of the Wheel? *American Economic Review,* 1994, 84(2), pp. 238–43.

STREETEN, P. Looking Ahead: Areas of Future Research in Human Development. *Journal of Human Development,* 2000, 1(1), pp. 25–48.

SUNDARAM, K., TENDULKAR, S. D. Poverty in India in the 1990s: An Analysis of Changes in 15 Major States, *Economic and Political Weekly,* 2003, 38(14), pp. 1385–1393.

SURYANARAYANA, M. H., AGRAWAL, A., PRABHU, S. *Inequality Adjusted Human Development Index for India's States.* New Delhi: UNDP, 2011.

UNDP. *Human Development Report. The Real Wealth of Nations: Pathways to Human Development.* New York: UNDP, 2010.

Recent Publications and Events

New Publications of the Czech Statistical Office

Agrocensus 2010 – Analytical Evaluation. Farm Structure Survey and Survey on Agricultural Production Methods. Prague: CZSO, 2012.

Demographic Yearbook of the Czech Republic 2011. Prague: CZSO, 2012.

DUBSKÁ, D., FERENC, J., KAMENICKÝ, J., MLÁDEK, T. *Tendence a faktory makroekonomického vývoje a kvality života v České republice v roce 2011* (Trends and Factors of the Macroeconomic Development and Quality of Life in the Czech Republic in 2011). Prague: CZSO, 2012.

Indicators of Social and Economic Development of the Czech Republic 1990 – 2nd Quarter 2012. Prague: CZSO, 2012.

External Trade of the Czech Republic in 2011. Prague: CZSO, 2012.

Historical Yearbook of National Accounts 1990–2010. Prague: CZSO, 2012.

Historie a současnost lesního a vodního hospodářství. Dlouhodobé časové řady (Past and Present of the Forestry and Water Economy. Long Time Series). Prague: CZSO, 2012.

Obyvatelstvo podle Sčítání lidu, domů a bytů 2011 – Česká republika (Population by the 2011 Population and Housing Census– Czech Republic). Prague: CZSO, 2012.

Vývoj obyvatelstva České republiky (2011) (Population Trends in the Czech Republic, 2011). Prague: CZSO, 2012.

Other Selected Publications

BÁRTA, V., BÁRTOVÁ, H. *Homo spotřebitel* (Homo Consumer). Prague: University of Economics, 2012.

CIPRA, T. *Penze: kvantitativní přístup* (Pensions: a Quantitative Approach). Prague: Ekopress, 2012.

EUROSTAT. *Eurostat Regional Yearbook 2012.* Luxembourg: Publications Office of the European Union, 2012.

EUROSTAT. *Taxation Trends in the European Union.* Luxembourg: Publications Office of the European Union, 2012.

KOZELSKÝ, T., VLACH, J. *Vývoj nákladů práce a jednotkových nákladů práce v ČR a EU na počátku krize* (The Development of the Labour Costs and the Unit Labour Costs in the Czech Republic and the EU at the Beginning of the Economic Crisis). Prague: VÚPSV, 2012.

LUKEŠ, M., JAKL, M. *Podnikání v České republice* (Business in the Czech Republic). Prague: University of Economics, 2012.

Health Care and Health Services in the Czech Republic 2011 in Statistical Data. Prague: Institute of Health Information and Statistics of the Czech Republic, 2012.

HORÁKOVÁ, M. *Mezinárodní pracovní migrace* v ČR (International Labour Migration in the Czech Republic). Prague: VÚPSV, 2011.

Main Economic and Social Indicators of the Czech Republic 1990–2011. Prague: VÚPSV, 2012.

OECD. *National Accounts of OECD Countries.* Paris: OECD, 2012.

OECD. *OECD Tourism Trends and Policies 2012.* Paris: OECD, 2012.

ROJÍČEK, M., ZAMRAZILOVÁ, E., SPĚVÁČEK, V., VINTROVÁ, R., ŽĎÁREK, V. *Makroekonomická analýza* (Macroeconomic Analysis). Prague: Linde, 2012.

UN. *2010 International Trade Statistics Yearbook.* New York: United Nations, 2012.
UN. *Statistical Yearbook 2010.* New York: United Nations, 2012.
UNIDO. *World Statistics on Mining and Utilities 2012.* Vienna: United Nations Industrial and Development Organization, 2012.

Conferences

The Czech Statistical Office (CZSO) had the honour to host the ***5th European Forum for Geostatistics (EFGS) 2012 Conference***, which took place ***from 24 to 26 October 2012*** in the Czech Association of Scientific and Technical Societies headquarters in the historical centre of ***Prague, Czech Republic***. The conference was organised by the European Forum for Geostatistics and the Czech Statistical Office. More information available at: *http://www.czso.cz/efgs/efgs2012.nsf/i/home.*

The Czech Statistical Office has organized also the ***International Marketing and Output Database Conference (IMAODBC) 2012***, which was held at the Conference Centre of the Průhonice Chateau, ***Prague, Czech Republic, from 8 to 12 October 2012***.

The ***New Techniques and Technologies for Statistics Conference (NTTS) 2013*** will take place ***from 5 to 7 March 2013 in Brussels, Belgium***, organised by Eurostat. More information available at: *http://www.ntts2013.eu.*

Information

Based on an agreement with the ***Czech Statistical Office the Ministry of Labour and Social Affairs of the Czech Republic*** replaces the present indicator of registered unemployment in the Czech Republic from January 2013 by a new one called ***Share of unemployed persons***, which expresses the share of unemployed job applicants aged 15 to 64 years of all residents of the same age. The new indicator is incomparable with the previous. More information available at: *http://portal.mpsv.cz/sz/stat, http://portal.mpsv.cz/sz/stat/nz/zmena_metodiky.*

Papers
We publish articles focused at theoretical and applied statistics, mathematics and statistical methods, econometrics, applied economics, economic, social and environmental analyses, economic indicators, social and environmental issues in terms of statistics or economics, and regional development issues.

The journal of *Statistika* has the following sections:
The *Analyses* section publishes high quality, complex, and advanced analyses based on the official statistics data focused on economic, environmental, and social spheres. Papers shall have up to 12,000 words or up to twenty (20) 1.5-spaced pages.

The *Methodology* section gives space for the discussion on potential approaches to the statistical description of social, economic, and environmental phenomena, development of indicators, estimation issues, etc. Papers shall have up to 12,000 words or up to twenty (20) 1.5-spaced pages.

The *Book Review* section brings reviews of recent books in the field of the official statistics. Reviews shall have up to 600 words or one (1) 1.5-spaced page.

Language
The submission language is English only. Authors are expected to refer to a native language speaker in case they are not sure of language quality of their papers.

Recommended Paper Structure
Title (e.g. On Laconic and Informative Titles) — Authors and Contacts — Abstract (max. 160 words) — Keywords (max. 6 words / phrases) — JEL classification code — Introduction — … — Conclusion — Annex — Acknowledgments — References — Tables and Figures

Authors and Contacts
Rudolf Novak*, Institution Name, Street, City, Country
Jonathan Davis, Institution Name, Street, City, Country
* Corresponding author: e-mail: rudolf.novak@domainname.cz, phone: (+420) 111 222 333

Main Text Format
Times 12 (main text), 1.5 spacing between lines. Page numbers in the lower right-hand corner. *Italics* can be used in the text if necessary. *Do not* use **bold** or underline in the text. Paper parts numbering: 1, 1.1, 1.2, etc.

Headings
1 FIRST-LEVEL HEADING (Times New Roman 12, bold)
1.1 Second-level heading (Times New Roman 12, bold)
1.1.1 Third-level heading (Times New Roman 12, bold italic)

Footnotes
Footnotes should be used sparingly. Do not use endnotes. Do not use footnotes for citing references (except headings).

References in the Text
Place reference in the text enclosing authors' names and the year of the reference, e.g. "White (2009) points out that…", "… recent literature (Atkinson et Black, 2010a, 2010b, 2011, Chase et al., 2011, pp. 12–14) conclude…". Note the use of alphabetical order. Include page numbers if appropriate.

List of References
Arrange list of references alphabetically. Use the following reference styles: [for a book] HICKS, J. *Value and Capital: An inquiry into some fundamental principles of economic theory.* Oxford: Clarendon Press, 1939. [for chapter in an edited book] DASGUPTA, P. et al. Intergenerational Equity, Social Discount Rates and Global Warming. In PORTNEY, P., WEYANT, J., eds. *Discounting and Intergenerational Equity.* Washington, D.C.: Resources for the Future, 1999. [for a journal] HRONOVÁ, S., HINDLS, R., ČABLA, A. Conjunctural Evolution of the Czech Economy. *Statistika, Economy and Statistics Journal,* 2011, 3 (September), pp. 4–17. [for an online source] CZECH COAL. *Annual Report and Financial Statement 2007* [online]. Prague: Czech Coal, 2008. [cit. 20.9.2008]. <http://www.czechcoal.cz/cs/ur/zprava/ur2007cz.pdf>.

Tables
Provide each table on a separate page. Indicate position of the table by placing in the text "insert Table 1 about here". Number tables in the order of appearance Table 1, Table 2, etc. Each table should be titled (e.g. Table 1 Self-explanatory title). Refer to tables using their numbers (e.g. see Table 1, Table A1 in the Annex). Try to break one large table into several smaller tables, whenever possible. Separate thousands with a *space* (e.g. 1 528 000) and decimal points with a *dot* (e.g. 1.0). Specify the data source below the tables.

Figures
Figure is any graphical object other than table. Attach each figure as a separate file. Indicate position of the figure by placing in the text "insert Figure 1 about here". Number figures in the order of appearance Figure 1, Figure 2, etc. Each figure should be titled (e.g. Figure 1 Self-explanatory title). Refer to figures using their numbers (e.g. see Figure 1, Figure A1 in the Annex).

Figures should be accompanied by the *.xls, *.xlsx table with the source data. Please provide cartograms in the vector format. Other graphic objects should be provided in *.tif, *.jpg, *.eps formats. Do not supply low-resolution files optimized for the screen use. Specify the source below the figures.

Formulas
Formulas should be prepared in formula editor in the same text format (Times 12) as the main text.

Paper Submission
Please email your papers in *.doc, *.docx or *.pdf formats to statistika.journal@czso.cz. All papers are subject to double-blind peer review procedure. You will be informed by our managing editor about all necessary details and terms.

Contacts
Journal of Statistika | Czech Statistical Office
Na padesátém 81 | 100 82 Prague 10 | Czech Republic
e-mail: statistika.journal@czso.cz
web: www.czso.cz/statistika_journal

Managing Editor: Jiří Novotný
phone: (+420) 274 054 299
fax: (+420) 274 052 133
e-mail: statistika.journal@czso.cz
web: www.czso.cz/statistika_journal
address: Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscription price (4 issues yearly)
CZK 369 (incl. postage) for the Czech Republic,
EUR 110 or USD 165 (incl. postage) for other countries.
Printed copies can be bought at the Publications Shop of the Czech Statistical Office (CZK 66 per copy).
address: Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscriptions and orders
MYRIS TRADE, s. r. o.
P. O. BOX 2 | 142 01 Prague 4 | Czech Republic
phone: (+420) 234 035 200,
fax: (+420) 234 035 207
e-mail: myris@myris.cz

Design: Toman Design
Layout: Ondřej Pazdera
Typesetting: Chráněná grafická dílna Slunečnice, David Hošek
Print: Czech Statistical Office

Published by the Czech Statistical Office
ISSN 1804-8765 (Online)
ISSN 0322-788X (Print)
Reg. MK CR E 4684