GRINCOH Growth-Innovation-Competitiveness Fostering Cohesion in Central and Eastern Europe

Identifying the drivers of inequality and poverty in the CEE EU Member States – a comparison with other EU countries applying the Shapley value approach to decomposition analysis

Sebastian Leitner and Robert Stehrer

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 290657.

Sebastian Leitner and Robert Stehrer
The Vienna Institute for International Economic Studies
WP5 Task1 P.5.1.
Final version: May 30, 2014

Identifying the drivers of inequality and poverty in the CEE EU Member States – a comparison with other EU countries applying the Shapley value approach to decomposition analysis

Abstract: Inequality is a multidimensional phenomenon though it is often discussed along a single dimension like income. This is also the case for the various decomposition approaches of inequality indices by recipients or income sources. In this paper we study one- and multidimensional indices on inequality on data for CEE EU Member States in comparison to other EU countries including four dimensions in our measure of multidimensional inequality: income, health, education and housing and apply various decomposition methods to these one- and multidimensional indices and also to a poverty index. In doing so, we apply standard decomposition techniques to the Mean logarithmic deviation index (I_0) and decompositions based on regression analysis in conjunction with the Shapley value approach to Gini indices.

1 Introduction

Inequality is a multidimensional phenomenon though it is often discussed along a single dimension like income which is the variable most often under consideration in this respect. This focus on a single variable - and income in particular - is even more the case when decompositions of inequality indices are applied. In this paper we instead consider inequality as a multidimensional concept for which different variables have to be taken into account simultaneously. Recently a large body of research started to focus on this multidimensional character of inequality together with the development of appropriate indices including more than one dimensions simultaneously (see Weymark, 2004; Justino, 2005; Lugo, 2005; Savaglio 2006a and 2006b; Cowell and Fiori, 2009). In this paper we provide a short discussion of the commonly suggested multidimensional indices on inequality and apply these using EU SILC data for CEE EU Member States and other EU countries (except for Cyprus and Ireland) for 2010. In doing so, we include four dimensions to study inequality: income, health, education and housing. This exercise yields important insights on how inequality (and the respective measures) changes when taking more dimensions of inequality into account.

The measurement of inequality along these dimensions and the cross-country comparisons is however only a first step. In the second step we contribute to explanations of these single- and multidimensional inequality indices by using decomposition methods (in line with the decomposition techniques known for one-dimensional decompositions methods with respect to income recipients). We apply various decomposition methods to these multidimensional indices: First, we apply standard decomposition techniques to the Mean logarithmic deviation index (I_0) – i.e. subgroup decompositions – and, second, a decomposition approach based on the Shapley value approach which allows to assess the relative importance of explanatory factors for inequality. The latter gained some attention in the one-dimensional case (see Shorrocks, 1999; Wan, 2004; Israeli, 2007; Molini and Wan, 2008; for example). To our knowledge this is the first attempt to apply this regression based technique to multidimensional inequality indices for a multitude of EU countries.

The paper goes as follows: In Section 2 we provide a brief discussion of important one- and multi-dimensional inequality indices used throughout in the paper. We then discuss the most important aspects of the data we use (sources, measurement issues, and definitions) in Section 3. Section 4 summarises some descriptive statistics on the data used, the results from the subgroup decomposition analysis to each of the four dimensions of inequality considered in this paper and the results from the subgroup decomposition for the multidimensional Mean logarithmic deviation index. In Section 5 we then introduce the concept of the Shapely decomposition and discuss the way we apply this method in the single- and multi-dimensional case. Further we present the results of this decomposition method for the one- and multidimensional case and a decomposition of the poverty headcount index for the one-dimensional case of household income. Section 6 concludes.

2. One- and multidimensional inequality

2.1 The one-dimensional case

Measuring and detecting the determinants of inequality based on household survey data has a long tradition in the literature. Already in the 1970s a wide range of inequality measures existed and their properties were described in detail e.g. in two essential publications of that strand of research, Sen's 'On Economic inequality' (1973) (see Sen, 1997) and Atkinson's 'The Economics of Inequality' (1975) (Atkinson, 1975). In general, inequality measurement is based on two different (classes of) measures, the first being the well-known and most frequently used Gini index,

$$G = \frac{N+1}{N-1} - \frac{2}{N(N-1)\mu} \sum_{i=1}^{N} \rho_i y_i$$

Here N denotes the number of observations, y_i is the variable under consideration (e.g. income) and ρ_i denotes the share of units with a specific income (or expenditure) value in the total population.[1] The second group of indices considered is the generalized class of entropy measures defined as

$$I_\alpha = \frac{1}{\alpha(\alpha-1)} \frac{1}{N} \sum_{i=1}^{N} \left[1 - \left(\frac{y_i}{\mu}\right)^\alpha\right] \text{ for } \alpha \neq 0{,}1$$

In both equations y_i denotes the income or expenditures (consumption) of the unit (individuals or households i), N is the number of units and μ is average income (or expenditure) in the total sample. In the formula of the generalized class of entropy measures, the parameter α can be seen as an indicator of inequality aversion and it also indicates the sensitivity to transfers at different parts of the distribution (for a negative α the index is sensitive to changes in the distribution that affect the lower tail); see Sen (1997) for a discussion and the frequently cited Jenkins (1995) for application and discussion. This allows, e.g., to focus on changes in the lower part of the income distribution, which might be more problematic with respect to social cohesion. For the limiting cases of $\alpha \rightarrow 0$ the entropy measure becomes Theil's second measure or the Mean logarithmic deviation

$$I_0 = \frac{1}{N} \sum_{i=1}^{N} ln \frac{\mu}{y_i}$$

which we apply in the single- and multidimensional case (see Section 4.2 below). For $\alpha \rightarrow 1$ it becomes the well-known Theil measure (I_1). For $\alpha = 2$ the measure becomes the half squared coefficient of variation I_2).

[1] Note that the Gini index can be expressed in different ways.

2.2 The multidimensional case

One of the first to introduce a measure of multidimensional distributions of well-being based on the theory of information was Maasoumi (1986, 1999); see also Lugo (2005) for a detailed discussion. He proposed to construct a multivariate inequality index in a two stage procedure. First, the attributes for each unit (e.g. individuals or households) are aggregated via an aggregator function yielding a real number S_i for each person. Second, a one-dimensional measure of inequality of the family of Generalised Entropy measures is calculated. This is based on the idea that different indicators of economic welfare are distributed differently; therefore Massoumi suggests an aggregator with a distribution that most closely represents the distributional information in each attribute. In particular he proposes a multivariate generalisation of the generalised entropy measure of divergence (the Kullback-Leibler distance) or closeness between the k densities (weighted sum of the pairwise divergence terms) and arrives at a distance measure D of the following form:

$$D_\beta(S, X, w) = \sum_{k=1}^{K} d_k \left\{ \sum_{i=1}^{N} S_i \left[\left(\frac{S_i}{x_{ik}} \right)^{-\beta} \right] / \beta(\beta - 1) \right\} \quad \text{for } \beta \neq 1$$

It is shown that the distribution of S which minimises D_β produces the optimal aggregation functions becomes

$$S_i = \left(\sum_{k=1}^{K} w_k x_{ik}^{\beta} \right)^{1/\beta}$$

where w_k is the weight given to the k-th attribute in the total aggregator function. The real number S_i denotes then the general weighted mean, called the 'well-being indicator for unit i, with the Cobb-Duglas function as special case (for $\beta \rightarrow 0$). The parameter β is related to the degree of substitutability between attributes and determines the shape of the contours for all pairs of attributes, identical for all pairs. The elasticity of substitution is given by $1/(1-\beta)$. The smaller β, the smaller is the elasticity of substitution between the attributes under consideration. For the second stage an index of the generalised entropy family is applied to the these weighted means S_i. In this paper we apply the index of Mean logarithmic deviation in Section 4.3, which in this case becomes (see also above)

$$I_{M0} = \frac{1}{N} \sum_{i=1}^{N} ln \frac{\mu}{S_i}$$

In Section 5 however, the results of the Shapley-value decomposition presented stem from a Gini index for income inequality and a multidimensional (Massoumi-based) Gini index.

3. Data

Data for the analysis presented in this paper is drawn from EU SILC for the year 2010. The four variables used as attributes for calculating the multidimensional welfare levels and the respective inequality indices thereof are: Household income, Household health status, Household education level and Housing level. Let us discuss them in turn.

The first dimension of multidimensional welfare considered is *household income*: In order to apply methodologically comparable household income data for all countries we adjusted the variable 'Total disposable household income' (HY020) by adding the variable 'Non-cash employee income' (PY020N). The resulting household income variable was then divided by the modified OECD equivalence scale (1-0.5-0.3), in order to obtain a household income variable adjusted for household composition differences. Obviously the needs of a household grow with each additional member but – due to economies of scale in consumption – not in a proportional way, e.g. for housing space, electricity, etc. With the help of equivalence scales each household type in the population is assigned a value in proportion to its needs. In our case a weight of 1 is assigned to the household head, a weight of 0.5 to all further members of the household aged 14 years or above and a weight of 0.3 to household members aged 0-13 years.

Household health status: For the analysis we used data on the subjective health status of all household members. We drew on the EU-SILC variable 'Genaral health' (PH010). The variable presents the subjective health status of a household member ranging from 1 (very good) to 5 (very bad). Since the health status of an individual obviously depends very much upon the age of the person, we calculated a 'conditional health status'. Thus we estimated for each individual country the linear age effect on subjective health with an OLS-regression (see Table 1 below) and used the estimation results to calculate a projected health status for every individual. The residual between the projected health status and the actual health status is taken as the 'conditional health status' of a person. The average of the 'conditional health status' over all household members is then used as the household health status. In addition we rescaled the variable from 0 to 1.

Household education level: For this indicator we use the mean level of years in education of all household members above 15 years of age who finished schooling or education in general. The years in education were calculated by using the variable highest education level attained by individuals (EU SILC variable "PE040: Highest ISCED level attained"). The household members were then assigned with the years in education needed to attain their respective education level. The household education level is than calculated as the average over those household members.

Housing level: Here we calculate a combined attribute from two variables: dwelling space and dwelling problems of the household. We used the number of rooms in the dwelling divided by the equivalised household size. The EU SILC (HH040, HH080/081, HH090/091 and HS160 to HS190) variables contain information on problems with the dwelling (e.g. not enough daylight, noise from neighbours or outside, etc.). For the variable dwelling problems we summed up the indicated problems for each household. Both variables dwelling space and dwelling problems were scaled from 0 to 1 and the mean of both taken to result in the final housing level of the respective household.

For the decomposition analysis by subgroups of the four above described attributes of the multidimensional inequality analysis in Section 4 and the Shapely value decomposition in Section 5 we used the following variables: gender and age group of the head of the household, urban versus rural household (EU SILC: DB100), educational attainment group (PE040) and activity status (PL031: employee, self-employed, unemployed, retired, others inactive) of the head of the household and household level employment rate (calculated as employed as a share of total household members above 15 years of age and not in education).

Table 1. OLS-regression results for subjective health status

	age		constant		R2
country	coefficient	p-value	coefficient	p-value	
AT	-0.027	0.000	5.177	0.000	0.232
BE	-0.021	0.000	4.881	0.000	0.164
BG	-0.035	0.000	5.306	0.000	0.440
CZ	-0.035	0.000	5.316	0.000	0.374
DE	-0.021	0.000	4.768	0.000	0.183
DK	-0.012	0.000	4.516	0.000	0.052
EE	-0.029	0.000	4.794	0.000	0.337
ES	-0.024	0.000	4.858	0.000	0.242
FI	-0.022	0.000	4.893	0.000	0.200
FR	-0.024	0.000	4.925	0.000	0.226
GR	-0.038	0.000	5.892	0.000	0.413
HU	-0.036	0.000	5.145	0.000	0.401
IT	-0.026	0.000	4.937	0.000	0.291
LT	-0.032	0.000	4.865	0.000	0.399
LU	-0.021	0.000	4.891	0.000	0.149
LV	-0.027	0.000	4.547	0.000	0.343
MT	-0.022	0.000	4.797	0.000	0.243
NL	-0.014	0.000	4.635	0.000	0.086
PL	-0.036	0.000	5.171	0.000	0.419
PT	-0.029	0.000	4.673	0.000	0.313
RO	-0.035	0.000	5.436	0.000	0.436
SE	-0.014	0.000	4.744	0.000	0.085
SI	-0.028	0.000	4.900	0.000	0.257
SK	-0.037	0.000	5.283	0.000	0.427
UK	-0.016	0.000	4.896	0.000	0.111

Source: EU-SILC 2010, own calculations.

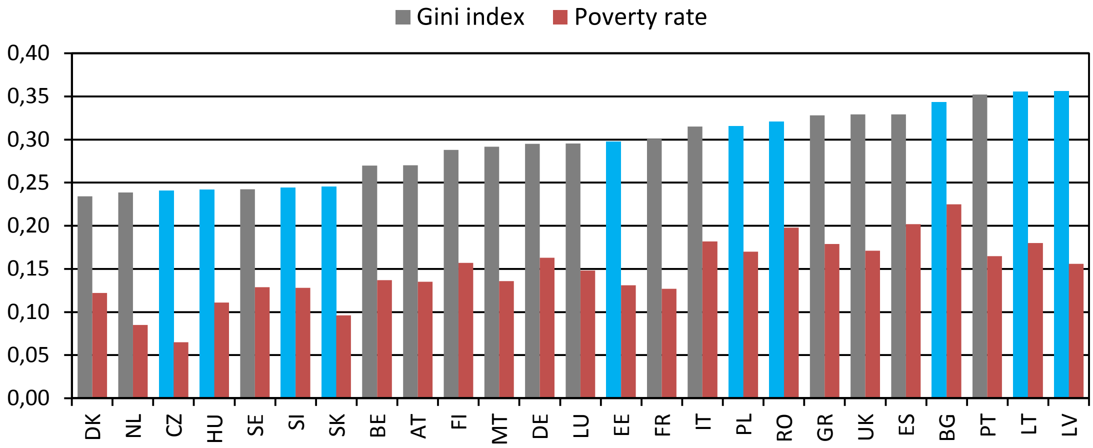
4 Descriptive results and subgroup decomposition

4.1 Descriptive results

In Figures 1 to 4 we present the Gini coefficients of one-dimensional inequality and poverty in the four attributes income, health status, housing and education for all EU countries (except for Cyprus and Ireland for which no EU SILC data is available for 2010). As can be seen, income inequality is, when measured by the Gini index, quite low in 2010 in most Central European and Scandinavian countries (see Figure 1) and much higher in e.g. Bulgaria, Lithuania and Latvia as well as the United Kingdom and the crisis-hit Southern EU rim (Spain, Portugal and Greece). The ranking of the countries doesn't change too much when ordered by poverty rates. Nevertheless, Bulgaria and Romania are those two CEE EU Member States that together with Spain account for the highest income poverty levels in the EU.

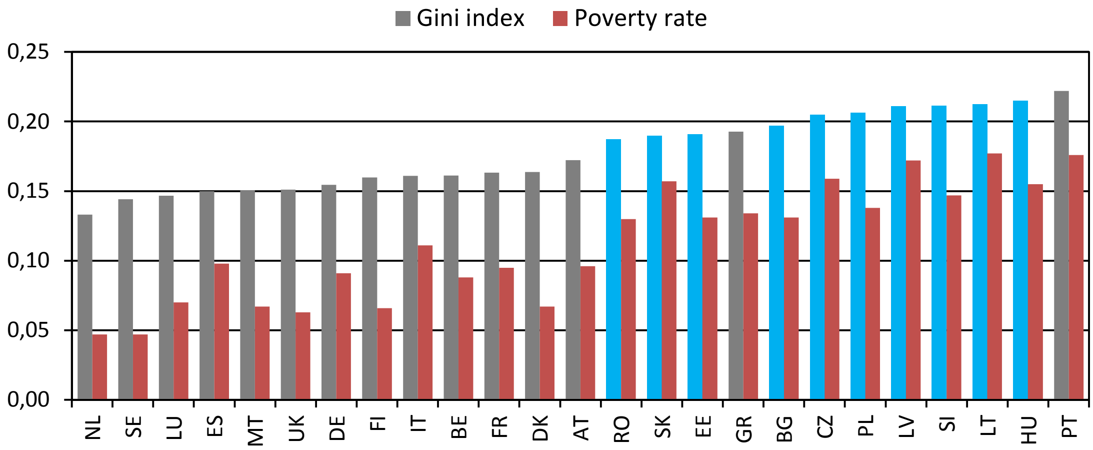
Obviously the inequality for the attributes conditional household health status and housing is lower than for the attribute income (see Figure 2 and 3). However, inequality in the conditional health status of households is in all CEE EU Member States (as in Portugal and Greece) higher compared to the rest of EU. Also concerning the level of inequality in housing quality (and space) a similar picture emerges. However, the inequality in housing in the Czech Republic, Slovenia and Slovakia again equal the Scandinavian countries as it was the case for the attribute income. Although inequality is in general much lower for housing compared to income, it should be pointed out that e.g. in Latvia the level of inequality in housing is almost double the one in most Scandinavian countries.

Figure 1. Gini indices and poverty rates of disposable household income p. c. equivalised



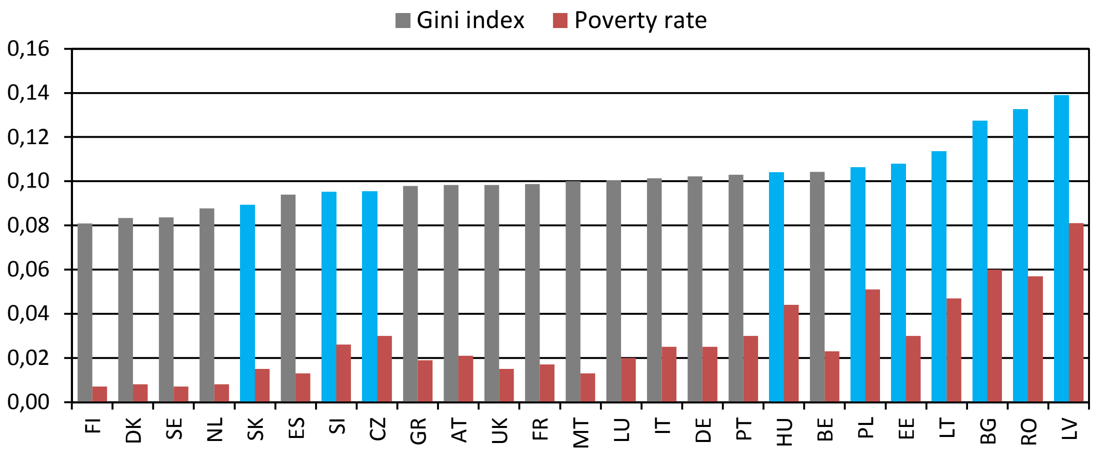
Source: EU SILC 2010, own calculations.

Figure 2. Gini indices and poverty rates of average conditional health status of households



Source: EU SILC 2010, own calculations.

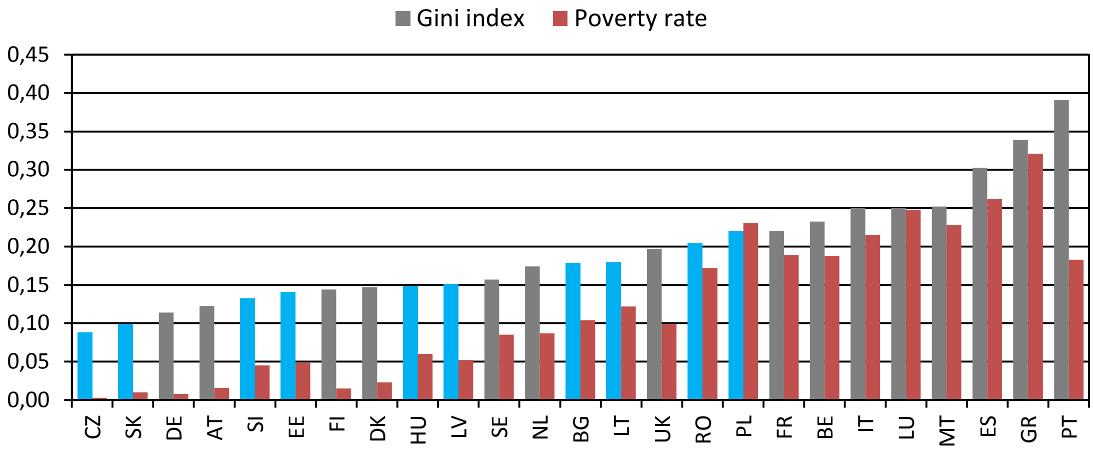
Figure 3. Gini indices and poverty rates of housing indicator



Source: EU SILC 2010, own calculations.

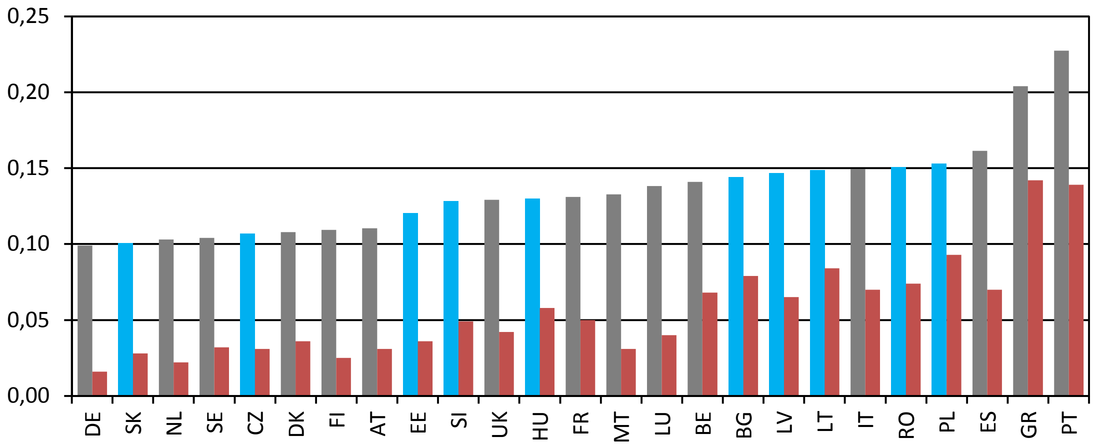
The average household education level however is also quite unequally distributed over the population (see Figure 4). Here, the level of inequality is much higher especially in South European countries compared to Central and North European countries. Remarkably, the CEE EU Member States have comparably low levels of inequality in this respect, although in Poland and Romania a relatively high share of households have substantially lower levels of education compared to the level of the median household in their country as shown by the respective poverty rates.

Figure 4. Gini indices and poverty rates of average educational attainment levels of households



Source: EU SILC 2010, own calculations.

Figure 5. Gini indices and poverty rates based on the Massoumi index of household multidimensional welfare



Source: EU SILC 2010, own calculations.

In order to study multidimensional inequality Figure 5 presents the Gini indices based on the Maasoumi index as discussed above. For the aggregation one has to specify a weight for each of the attributes considered. We applied the same weights to the four above described attributes[2] which we scaled from 0 to 1. Another choice has to be made on the degree of substitutability in the aggregation function. The higher the elasticity of substitution the lower is the level of the multidimensional inequality index. A higher degree of substitutability means that low levels on one of the attributes can be compensated more easily by high levels on another (Lugo, 2005). For Figure 5 the presented Gini indices where based on a Massoumi index for $\beta = 0.25$. This value was chosen in order to present results for a case where some, but not perfect substitution is possible.

As can be seen from Figure 5 the results of the Gini indices with respect to the ranking of countries is very much driven by the attributes that have the highest levels in inequality, which are the income inequality and the inequality in educational attainment levels of households. Thus, again Slovakia and the Czech Republic are to be found in the group of EU countries with low levels of inequality, Bulgaria, Romania, Poland and the two Baltic countries Latvia and Lithuania among those with comparably high levels and Slovenia, Hungary and Estonia in the middle group.

4.2 Subgroup decomposition

In this section we present results from the decomposition analysis based on the Mean logarithmic deviation as discussed above. The decomposition of the Mean logarithmic deviation (MLD) inequality index can be applied in the one-dimensional case as well as in the multi-dimensional case for an analysis of the determinants of inequality observed by income recipients. The MLD can be decomposed in two terms, the within and the between component

$$I_0 = \sum\nolimits_k v_k I_{0,k} + \sum\nolimits_k v_k \ln(1/\lambda_k)$$

where v_k denotes population shares and $\lambda_k = \mu_k/\mu$. The first term, the within component of the MLD, represents the part of the total inequality that is due to variations within the population subgroups, whereas the between component represents the part of the total inequality that accrues from differences between the means of the population subgroups.

In Tables 2-6 we present the results of the decomposition into between and within group effects of the individual attributes of the multidimensional inequality indicator as well as the results of the decomposition when applying it to the Massoumi index for $\beta = 0.25$. The decomposition of inequality is performed by population subgroups according to different characteristics of the household or the head of the household observed. These are the gender of the head of household, the age group, urban versus rural location, education level and employment status of the head of the household as well as the household level employment rate (calculated as employed as a share of total household members above 15 years of age not being in education anymore).

The higher the between component as a share of the total inequality index, in our case the Mean logarithmic deviation (I_0), the more the respective characteristic can be seen as source of inequality in an attribute. However, the magnitude of the within and between component also depends on the partition of the population into subgroups. The higher the number of subgroups which are considered in the decomposition analysis of a specific characteristic, the higher the between group component will become

[2] Changing the weight of an attribute obviously raises or lowers the Massoumi index depending upon if the level of inequality of the attribute is higher or lower than that of the overall Massoumi index. A change of weights however does not alter the structure of the below presented results of the decomposition analysis, only the magnitude of the results change.

by definition. Therefore the results of the decomposition analysis into within and between group components should be interpreted cautiously. Comparisons over time or cross country with the same number of subgroups however can be interpreted without difficulty. In this paper we compare the results of the decomposition analysis for all EU countries (except for Cyprus and Ireland for which no data is available for that year) for 2010 in a cross-country perspective. In our analysis the number of dimensions in each subgroup does not differ too much, such that also a comparison across dimensions is done, though with care.

4.2.1 Decomposition of equivalised per capita household income

As can be seen from Table 2 the results for the EU countries differ quite substantially concerning the characteristics of heads of households influencing household income levels. In about half of the countries analysed household income is most strongly influenced by the education level of the head of household. On average of all countries about 16% of the income differences are accrued from that characteristic. About 15% of the income differences can be explained by the employment rates of the individual households. Further, also the characteristic of the employment status of the head of the household explains with about 10% quite a large part of income differentiation. Differences between the mean income levels of the seven age groups of the heads of household account for only about 5% of the total Mean logarithmic deviation (I_0). Much lower differences can be detected when looking at the decomposition by rural and urban households (accounting for about 3% of income differences). However, in Bulgaria, Hungary, Poland and Romania alike, households residing at the countryside face a pronounced worse income positions than urban households. Only in the Czech Republic households headed by women face a substantially lower income position in general than those headed by men.

Table 2. Multidimensional inequality decomposition: Attribute household income
Between group components as % of Mean logarithmic deviation (I_0)

Decomposition by	**AT**	**BE**	**BG**	**CZ**	**DE**	**DK**	**EE**
gender	1.18	2.27	0.36	6.03	1.42	0.05	1.44
age	4.20	5.11	9.62	9.15	3.01	15.39	6.06
urban / rural regions	0.45	0.01	10.96	1.57	0.21	0.61	2.52
education	12.74	18.15	24.33	16.36	11.79	9.31	11.27
empl. status	8.34	14.31	14.81	19.35	12.91	10.66	15.07
hh-empl-rate	10.85	19.71	25.47	26.67	12.01	19.54	25.14
Decomposition by	**ES**	**FI**	**FR**	**GR**	**HU**	**IT**	**LT**
gender	0.27	0.24	1.12	1.90	1.60	0.64	0.55
age	1.96	10.78	5.44	3.67	0.92	2.35	1.05
urban / rural regions	2.00	2.64	.	1.61	6.80	0.57	2.39
education	10.79	14.95	12.46	16.48	22.48	9.73	9.10
empl. status	6.88	12.65	4.14	3.42	9.98	5.55	8.55
hh-empl-rate	9.04	18.78	2.20	7.23	13.01	8.11	10.44
Decomposition by	**LU**	**LV**	**MT**	**NL**	**PL**	**PT**	**RO**
gender	0.67	0.29	1.10	1.05	0.67	1.33	2.26
age	1.72	5.29	3.54	5.30	1.53	3.52	2.65
urban / rural regions	0.59	2.39	.	.	7.12	3.33	14.12
education	21.50	14.60	12.02	13.34	20.42	33.70	32.68
empl. status	7.14	10.96	11.22	6.00	9.41	7.32	11.64
hh-empl-rate	7.14	16.61	25.22	8.59	12.12	7.93	6.14
Decomposition by	**SE**	**SI**	**SK**	**UK**			

gender	0.21	0.24	1.88	1.62
age	10.88	3.26	6.60	6.06
urban / rural regions	1.07	.	3.02	0.44
education	5.78	23.54	9.39	13.85
empl. status	6.38	10.74	13.37	17.30
hh-empl-rate	13.39	17.56	24.08	20.25

Source: EU-SILC 2010, own calculations.

4.2.2 Decomposition of household health status

The decomposition of inequality of the aggregated health status of households showed that subjective health is obviously strongly influenced by the age characteristic of the household head. Since this fact may distort also other decomposition results, we calculated a conditional health variable, being the divergence of subjective health from a health status projected according to age, as already discussed above. This conditional health status was rescaled to 0 to 1. As we know from Figure 2 the conditional health status is quite equally distributed across households in all countries. Moreover, the health status of households varies only very little by gender of the head of the household and between urban and rural regions in all countries.

Table 3. Multidimensional inequality decomposition: Attribute household health status
Between group components as % of Mean logarithmic deviation (I_0)

Decomposition by	AT	BE	BG	CZ	DE	DK	EE
gender	0.34	0.59	0.15	0.62	0.02	0.01	0.35
age	8.17	4.87	10.01	9.96	6.57	1.11	11.16
urban / rural regions	0.00	0.02	0.44	0.02	0.03	0.03	0.00
education	3.50	3.45	2.54	2.36	1.15	1.08	2.66
empl. status	7.25	6.19	7.99	7.98	6.59	1.90	9.97
hh-empl-rate	7.04	6.20	9.64	7.77	6.67	1.53	7.59

Decomposition by	ES	FI	FR	GR	HU	IT	LT
gender	0.23	0.00	0.42	1.20	0.70	0.52	0.14
age	8.68	7.88	9.13	13.03	10.02	10.27	9.96
urban / rural regions	0.04	0.24	0.05	0.32	0.16	0.03	0.12
education	5.97	2.89	5.47	7.31	3.51	5.87	5.46
empl. status	7.08	5.20	6.79	7.45	7.92	6.58	8.80
hh-empl-rate	7.20	3.78	6.80	8.04	8.46	8.02	8.21

Decomposition by	LU	LV	MT	NL	PL	PT	RO
gender	0.25	0.19	1.59	0.27	0.55	0.36	2.01
age	5.14	7.38	12.38	3.37	11.82	9.04	11.83
urban / rural regions	0.13	0.18	0.09	.	0.06	0.17	0.06
education	3.50	2.78	7.90	2.29	3.50	5.86	4.38
empl. status	6.03	6.57	9.40	3.68	8.74	7.58	9.84
hh-empl-rate	6.11	6.25	10.32	3.71	10.05	7.75	10.76

Decomposition by	SE	SI	SK	UK
gender	0.19	0.14	0.95	0.12
age	2.71	4.30	11.97	3.70
urban / rural regions	0.06	.	0.25	0.17
education	1.78	1.98	3.88	2.80
empl. status	2.50	3.72	10.24	5.53
hh-empl-rate	2.33	3.66	11.89	5.93

Source: EU-SILC 2010, own calculations.

Nevertheless, as can be seen from Table 3 below, in all countries the education level of the head of household seems to have some influence also on the health status of the respective household. Also those households with higher household employment rates and those headed by employed persons face a better health status.

4.2.3 Decomposition of household education level

Decomposing household education levels by the age group of the head of household indicates that obviously younger age cohorts had the chance to attain higher education levels (see Table 4). However, in most South European and some transition countries, i.e. Spain, Greece, Hungary, Italy, Malta, Lithuania, Slovakia, Romania but also France and Sweden the differences in education between younger and older age cohorts are much more pronounced. Moreover, in Bulgaria, Hungary, Slovakia and Romania there are substantially higher differences between urban and rural households than in other EU countries. The decomposition by level of education of the head of households shows that educational segregation differs quite a lot in EU countries. It is particularly high in the Czech Republic, Germany and Hungary. This also means that the level of formal education of children in those countries strongly depends upon the educational level attained by their parents. In France, Hungary, Malta, Sweden and the Slovak Republic households with higher employment levels also have substantially higher aggregate education levels.

Table 4. Multidimensional inequality decomposition: Attribute household education level
Between group components as % of Mean logarithmic deviation (I_0)

Decomposition by	AT	BE	BG	CZ	DE	DK	EE
gender	0.28	0.21	0.04	2.19	0.64	0.00	0.19
age	2.50	4.09	2.39	6.17	3.65	3.03	6.08
urban / rural regions	0.09	0.01	4.66	2.04	0.00	0.00	2.14
education	36.36	35.03	43.15	76.67	78.64	33.89	52.66
empl. status	3.39	5.17	3.28	8.91	5.69	4.59	7.01
hh-empl-rate	4.41	5.83	4.66	9.53	4.86	6.80	7.96

Decomposition by	ES	FI	FR	GR	HU	IT	LT
gender	0.06	0.00	0.24	0.68	1.26	0.12	0.07
age	9.60	5.64	12.66	7.35	8.75	9.48	8.67
urban / rural regions	1.17	0.44	0.62	1.31	4.99	0.45	2.18
education	33.90	37.84	54.45	28.89	77.54	33.38	41.90
empl. status	8.79	5.70	11.48	5.78	8.47	6.90	7.96
hh-empl-rate	9.39	9.33	11.88	7.46	11.03	8.62	8.42

Decomposition by	LU	LV	MT	NL	PL	PT	RO
gender	0.17	0.00	1.08	0.33	0.16	0.11	1.92
age	3.31	3.76	10.77	6.18	5.04	6.59	8.84
urban / rural regions	0.22	1.74	0.00	.	1.78	0.87	5.83
education	65.66	40.68	42.37	52.57	35.06	27.44	41.30
empl. status	5.17	4.29	9.60	8.55	5.91	5.15	7.11
hh-empl-rate	5.26	5.33	11.86	8.10	7.10	5.82	5.55

Decomposition by	SE	SI	SK	UK
gender	0.01	0.03	1.16	0.02
age	7.89	3.86	9.05	1.24
urban / rural regions	0.53	.	3.45	0.11
education	48.97	50.74	57.78	22.01
empl. status	7.79	4.61	9.02	1.50
hh-empl-rate	12.41	4.72	10.57	1.49

Source: EU-SILC 2010, own calculations.

4.2.4 Decomposition of housing level

The data underlying the fourth attribute, housing level (quality and space p.c.), shows quite low differentiation between households in general (see Figure 3 above). From Table 5 we can see that the characteristics used in the decomposition analysis do not give much insight into the existing inequality with respect to housing. In some countries, especially North European countries but also France, older age cohorts seem to face higher levels of housing most probably due to more dwelling space. Furthermore in Bulgaria, Estonia, Greece, Lithuania and Portugal the level of housing of households in urban areas is lower than that of rural households. This result is obviously driven by less living space of dwellings in urban areas. Only in Romania and Bulgaria the housing quality is markedly distinct between education levels. However, here the higher the education level the lower the floor space of dwellings on average, since e.g. people with tertiary education most probably live in urban areas. Only in some EU countries those households being more active on the labour market face better housing levels (according to the constructed index).

Table 5. Multidimensional inequality decomposition: Attribute housing
Between group components as % of Mean logarithmic deviation (I_0)

Decomposition by	AT	BE	BG	CZ	DE	DK	EE
gender	0.03	0.77	0.00	0.15	0.11	0.21	0.11
age	3.72	8.15	1.84	2.84	9.06	12.65	0.65
urban / rural regions	7.04	3.74	0.47	2.67	3.93	2.99	0.01
education	0.35	0.94	4.24	0.04	0.93	0.29	3.05
empl. status	2.72	7.25	2.15	1.99	7.43	7.16	2.18
hh-empl-rate	2.60	7.09	2.23	2.98	1.86	5.89	3.49

Decomposition by	ES	FI	FR	GR	HU	IT	LT
gender	1.05	0.03	0.31	1.30	0.12	1.02	0.01
age	7.96	9.78	10.59	2.67	2.01	3.01	0.65
urban / rural regions	0.32	2.62	5.23	4.27	0.00	2.21	0.04
education	0.93	1.03	1.14	0.87	2.02	1.02	2.22
empl. status	4.34	5.05	8.04	1.76	2.94	2.11	1.43
hh-empl-rate	15.09	2.47	6.23	4.23	1.69	6.70	2.02

Decomposition by	LU	LV	MT	NL	PL	PT	RO
gender	0.44	0.00	2.48	0.64	0.06	1.04	0.18
age	10.50	0.50	4.64	9.50	0.87	3.90	0.42
urban / rural regions	4.01	0.03	0.39	.	0.16	3.65	1.52
education	1.00	2.39	1.27	0.39	2.53	0.72	6.31
empl. status	7.68	0.99	1.96	3.92	1.47	3.16	3.00
hh-empl-rate	10.06	1.49	11.15	10.74	1.20	5.82	1.08

Decomposition by	SE	SI	SK	UK
gender	0.00	0.00	0.00	0.22
age	10.80	0.57	4.10	11.06
urban / rural regions	2.49	.	1.10	1.11
education	0.62	1.45	0.00	0.94
empl. status	5.58	0.89	2.93	9.51
hh-empl-rate	2.17	0.81	2.52	7.13

Source: EU-SILC 2010, own calculations.

4.3 Decomposition of multidimensional inequality

Finally, we present the results for the decomposition of the multidimensional index in Table 6. As already mentioned above, all attributes considered (equivalised per capita household income, the mean of the conditional health status of all household members, the mean of the education levels of household members and the housing indicator) are given the same weights. The parameter β is set 0.25[3], which offers a medium substitutability between the four attributes. The Mean logarithmic deviation (I_0) inequality index was then calculated and decomposed by the respective characteristics of the head of the household and the household characteristics as reported in Table 6.

Table 6. Multidimensional inequality decomposition: Massoumi inequality index ($\beta = 0.25$)
Between group components as % of Mean logarithmic deviation (I_0)

Decomposition by	AT	BE	BG	CZ	DE	DK	EE
gender	1.49	1.44	0.23	3.07	0.96	0.15	1.17
age	7.66	7.89	12.37	15.29	3.87	7.28	14.63
urban / rural regions	0.12	0.43	4.04	0.00	0.10	0.00	0.74
education	32.08	48.15	42.38	19.37	28.09	29.84	30.38
empl. status	15.69	16.89	13.03	19.19	21.22	20.97	21.96
hh-empl-rate	19.33	18.73	20.64	20.80	16.81	20.59	26.06

Decomposition by	ES	FI	FR	GR	HU	IT	LT
gender	0.34	0.00	1.21	4.21	2.43	1.14	0.48
age	16.36	13.71	11.72	21.20	14.52	18.30	16.91
urban / rural regions	1.78	0.57	0.00	1.14	2.01	0.27	2.05
education	50.66	43.37	46.78	52.33	33.83	49.17	41.29
empl. status	17.77	26.71	14.93	18.16	16.04	15.51	19.43
hh-empl-rate	21.32	34.82	16.55	26.52	22.14	22.62	23.45

Decomposition by	LU	LV	MT	NL	PL	PT	RO
gender	0.56	0.29	2.44	1.02	1.29	1.25	4.51
age	2.19	11.61	19.13	7.40	15.74	18.89	18.80
urban / rural regions	1.40	1.66	0.01	.	1.46	1.23	6.43
education	53.08	30.60	52.90	42.13	34.57	52.47	50.90
empl. status	9.17	14.77	19.86	17.05	16.97	14.89	17.95
hh-empl-rate	8.83	19.57	25.96	15.45	22.14	18.12	17.50

Decomposition by	SE	SI	SK	UK
gender	0.32	0.33	3.29	0.41
age	13.59	8.22	21.75	3.85
urban / rural regions	0.00	.	2.01	0.87
education	32.85	24.64	31.17	33.05
empl. status	17.14	10.09	21.57	12.77
hh-empl-rate	26.63	12.33	33.60	12.71

Source: EU-SILC 2010, own calculations.

As we can see from Table 6 and the findings above the decomposition results of multidimensional inequality are strongly driven by those attributes with the highest inequality levels, which in our case are the household income and the household education level. Hence, in the case of the multitude of countries analysed, the multidimensional welfare levels of households are mostly influenced by the differentiation

[3] A lower value of ß would obviously raise the value of the inequality index.

with respect to the education level of the head of household. On average, differences between education groups account for almost 40% of total differences in household welfare variations. The employment status of the head of the household (on average 17%) and the level of labour market activity of household members (on average 21%) also exert strong influence on the level of well-being, less so the age of the head of household (13%). The other characteristics of the households analysed, i.e. gender (1.4%) and rural versus urban households (1.2%) have only minor or negligible effects. Somewhat stronger differences in multidimensional welfare levels between urban areas and the countryside are only evident in Romania (6.4%) and Bulgaria (4%). However, the explanatory value of the characteristics of heads of households and household members used in the decomposition analysis vary quite large between different EU countries.

5. A Shapley-value decomposition of multidimensional inequality indices

5.1 Outline of decomposition procedure

In this section we undertake a decomposition analysis based on regression analysis, the Shapley value approach described below. [4] To our knowledge such a regression based approach to one- and multidimensional inequality decomposition was not yet undertaken in the literature at least for a multitude of EU countries. Compared to the subgroup decomposition approach undertaken in Section 4 the advantage of a regression based approach is that the relative importance of many variables and groups of them to explain inequality (like age, gender, educational attainment, employment status etc.) are taken into account simultaneously. Thus, the regression approach (step 1) allows assessing the importance of each of these explanatory variables conditional on all other variables for each of the respective dimension of inequality considered (income, health, education and housing). The Shapely value approach (step 2) then further allows calculating the contribution of each of these explanatory variables to the respective inequality measure and via the aggregator function as outlined above also to the multidimensional inequality measure.

Recently, the literature on inequality analysis has provided various decomposition methods which are based on regression results like the Shapley value approach as introduced by Shorrocks (1999) but also others; see Fields and Yoo (2000), Morduch and Sicular (2002), Fields (2003), Wan (2004), Gunatilaka and Chotikapanich (2006) or Molini and Wan (2008) for such applications; see also Cowell and Fiorio (2009) for a critical review. Finally, Israeli (2007) shows how the Shapley approach is related to the method proposed by Fields (2003) and also points to some advantages of the former which is applied in this paper. The most important advantage of the Shapley value approach is that this takes the potential correlation amongst regressors into account. One should however note that the latter two contributions aim at decomposing the R2, i.e. the explained part of the regressions, whereas in this paper we decompose the resulting inequality measure. This is more similar to the contributions by Wan (2004), Gunatilaka and Chotikapanich (2006) and Molini and Wan (2008).

The Shapley value approach can be illustrated by using a simple example with three explanatory variables. We first regress individual income levels y on these explanatory variables x_i (i = 1, 2, 3),

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \,,$$

where ε denotes the error term. The predicted income level is then given by

4 The authors already applied the approach in the one-dimensional case to Western Balkan countries: see Leitner and Stehrer (2009).

$\hat{y}_{123} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$.

This predicted value is then used to calculate the Gini coefficient $\hat{G}^{(0)}_{\{123\}}$, where subscripts denote the variables included. In the first round we then eliminate one variable and calculate the predicted income levels $\hat{y}_{\{23\}}, \hat{y}_{\{13\}}$ and $\hat{y}_{\{12\}}$. The corresponding Gini coefficients are then given by $\hat{G}^{(1)}_{\{23\}}, \hat{G}^{(1)}_{\{13\}}$ and $\hat{G}^{(1)}_{\{12\}}$ respectively. Analogously, in the second round we eliminate two variables, thus calculating $\hat{y}_{\{1\}}, \hat{y}_{\{2\}}$ and $\hat{y}_{\{3\}}$. The resulting Gini coefficients are $\hat{G}^{(2)}_{\{1\}}, \hat{G}^{(2)}_{\{2\}}$ and $\hat{G}^{(2)}_{\{3\}}$. The final round would then be to include the constant only; the resulting Gini coefficient would thus be $\hat{G}^{(3)}_{\{\ \}} = 0$.

The marginal contributions are then calculated using the Gini coefficients. The first round marginal contributions for each variable are $C_1^{(1)} = \hat{G}^{(0)}_{\{123\}} - \hat{G}^{(1)}_{\{23\}}$, $C_2^{(1)} = \hat{G}^{(0)}_{\{123\}} - \hat{G}^{(1)}_{\{13\}}$ and $C_3^{(1)} = \hat{G}^{(0)}_{\{123\}} - \hat{G}^{(1)}_{\{12\}}$. The marginal contributions in the second round of the first variable are given by

$$C_1^{(2,1)} = \hat{G}^{(1)}_{\{12\}} - \hat{G}^{(2)}_{\{2\}} \text{ and } C_1^{(2,2)} = \hat{G}^{(1)}_{\{13\}} - \hat{G}^{(2)}_{\{3\}}$$

The average of these contributions is the marginal contribution of the first variable in the second round, i.e. $C_1^{(2)} = \frac{1}{2}\left(C_1^{(2,1)} + C_1^{(2,2)}\right)$. Similarly we calculate $C_2^{(2)}$ and $C_3^{(2)}$. The third round contribution is given by $C_1^{(3)} = \hat{G}^{(2)}_{\{1\}} - \hat{G}^{(3)}_{\{\ \}} = \hat{G}^{(2)}_{\{1\}}$ as $\hat{G}^{(3)}_{\{\ \}} = 0$ and analogously for $C_2^{(3)} = \hat{G}^{(2)}_{\{2\}}$ and $C_3^{(3)} = \hat{G}^{(2)}_{\{3\}}$.

Finally, averaging the marginal contributions of each variable over all rounds results in the total marginal effect of each variable $j = 1,2,3$ i.e.

$$C_j = \frac{1}{3} \cdot \left(C_j^{(1)} + C_j^{(2)} + C_j^{(3)}\right).$$

The proportion of inequality not explained is then given by

$$C_R = G - \hat{G}^{(0)}_{\{123\}}.$$

The approach can easily be extended to any number of explanatory factors and to other inequality measures. The Shapley value method can furthermore be applied to the decomposition of a poverty indicator, as will also be done in Section 5.

Wan (2002) points to the fact that the presence of a negative constant in the regression equation may lead to negative predicted individual income levels. In that case the calculation of a Gini-coefficient and thus the contributions of individual variables to overall inequality would be impossible. To overcome this pitfall he shows in Wan (2004) that different model specifications can be used for the underlying estimated income generating function, delivering moreover better log-likelihood values than the linear estimation model. Following his approach, we choose for the analysis in this paper a semilog model:

$\ln y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$,

Since we are not interested in the decomposition of the log of income, but income we have to take the antilog of the above model resulting in

$e^{\ln y} = e^{\beta_0} * \left(e^{\beta_1}\right)^{x_1} * \left(e^{\beta_2}\right)^{x_2} * \left(e^{\beta_3}\right)^{x_3} * e^{\varepsilon}$, which is

$$y = e^{\beta_0} * \left(e^{\beta_1}\right)^{x_1} * \left(e^{\beta_2}\right)^{x_2} * \left(e^{\beta_3}\right)^{x_3} * e^{\varepsilon} .$$

The simple advantage of this model is that the constant e^{β_0} is now a positive scalar, which does not influence the magnitude of the calculated Gini coefficient. The elimination procedure as described above however remains unchanged.

5.2 Empirical results

5.2.1 Step 1: Regression analysis

Following the above described approach we first regress the logarithm of equivalised household income on the variables age, gender, employment status and highest education level attained of the head of the household as well as the calculated employment share of the household and a regional dummy. Since for the latter variable no data was available for the Netherlands, Malta and Slovenia (apart from Ireland and Cyprus for which no EU SILC data for 2010 is available at all) we had to drop those countries for the analysis. The results are reported in Table 7 below. The explained part of the variance is on average slightly above 24% as shown by R2. In general the coefficients of the variables applied have the expected signs and are significant. Coefficients for Age and Age2 show that household incomes rise with increasing age of the head of the household. Households headed by men accrue higher incomes and higher education levels are correlated with higher income levels. The same significant result is to be found for the variable household employment share. The dummy variables for employment status of head of household (differences in income to the employment status unemployed) deliver the expected signs in almost all countries; however the results are not always significant. The regression results for the multidimensional case are moved to the Appendix Table A.1. The results are similar to those with respect to household income (Coefficient signs remain in general the same, whilst the share of the explained variance rises to about 40% on average). The final output of the Shapely decomposition of multidimensional inequality is discussed in the next section.

Table 7. Regression results: Dependent variable: Logarithm of equivalised household income (modified OECD equivalence scale)

		AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GR
Age		0.0255***	0.0235***	0.0075*	0.0101***	0.0129***	0.0555***	-0.0159***	0.0332***	0.0409***	0.0425***	0.0449***
$Age^2/100$		-0.0172***	-0.0208***	-0.0083**	-0.0079***	-0.0080***	-0.0500***	0.0139**	-0.0216***	-0.0348***	-0.0296***	-0.0320***
Male		0.0192	0.1105***	0.0909***	0.1504***	0.0877***	0.0049	0.1086***	0.0509	0.0453***	0.1139***	0.1462***
Education level	Upper secondary	0.2682***	0.1171***	0.2895***	0.1071***	0.1241***	0.1140***	0.0395	0.2685***	0.1201***	0.1763***	0.3248***
	Tertiary	0.4549***	0.3625***	0.5514***	0.3818***	0.3648***	0.2771***	0.2919***	0.4814***	0.3935***	0.5079***	0.6488***
Household employment share		0.4669***	0.5703***	0.8307***	0.5267***	0.4798***	0.4713***	0.7492***	0.9388***	0.5501***	0.4017***	0.6174***
Rural region		-0.0316**	0.0399**	-0.1836***	-0.0575***	-0.0345***	-0.0382***	-0.0813***	-0.1080***	-0.1131***	-0.0278**	-0.0805**
Employment status	Employee	0.0491	-0.1103**	0.1922***	0.1578***	0.2060***	-0.0806**	0.0596	0.1519*	0.1000***	-0.0058	0.3801***
	Self-employed	0.1930***	0.1502***	0.1712***	0.1917***	0.3980***	-0.0359	0.2726***	0.4450***	0.2061***	0.1339***	0.3697***
	Retired	0.1206**	-0.0499	0.1292**	0.2084***	0.4709***	-0.1617***	-0.0184	-0.4722***	0.1042***	0.1674***	0.2176**
	Other inactive	0.2225***	0.2850***	0.4050***	0.2534***	0.4390***	0.0708*	0.3313***	0.6023***	0.3136***	0.2849***	0.5594***
Constant		8.4782***	8.5921***	7.0189***	7.9171***	8.5398***	8.4705***	8.3514***	7.3473***	8.3156***	7.9956***	6.9154***
R2		0.2613	0.2274	0.3717	0.3629	0.2447	0.3146	0.1940	0.0885	0.2300	0.2215	0.1050
Observations		6187	6125	6161	9096	12900	5804	4963	13500	11000	11000	6967

		HU	IT	LT	LU	LV	PL	PT	RO	SE	SK	UK
Age		0.0043*	0.0465***	-0.0179	0.0069*	-0.0105	0.0075**	0.0307***	0.0020	0.0598***	0.0085**	0.0112***
$Age^2/100$		-0.0006	-0.0262***	0.0280*	0.0039	0.0059	-0.0013	-0.0268***	0.0020	-0.0524***	-0.0073**	-0.0087**
Male		0.1007***	0.1366***	0.0726	-0.0001	-0.0178	0.1099***	0.1532***	0.1334***	0.0003	0.0896***	0.0818***
Education level	Upper secondary	0.1873***	0.2739***	0.3344***	0.2004***	0.1671***	0.1784***	0.4724***	0.3431***	0.1262***	0.1067***	0.0986***
	Tertiary	0.4981***	0.6028***	0.6154***	0.5568***	0.5924***	0.6075***	0.9186***	0.8669***	0.2676***	0.3074***	0.3233***
Household employment share		0.5073***	0.9787***	1.0346***	0.4758***	0.8550***	0.5388***	0.6629***	0.5466***	0.4209***	0.7717***	0.5559***
Rural region		-0.1138***	-0.0130	-0.1376**	0.0501***	-0.1159***	-0.1775***	-0.1233***	-0.2426***	-0.0685**	-0.0868***	0.0235
Employment status	Employee	0.1503***	0.9221***	0.0481	0.1474***	0.4854***	0.2279***	0.0696	0.1252*	0.0333	0.3524***	0.3075***
	Self-employed	0.2179***	1.1504***	0.7243***	0.2197***	0.6245***	0.3285***	0.1005**	0.4294***	0.1973***	0.3139***	0.5082***
	Retired	0.1773***	1.0590***	0.6780***	0.0746	0.3185**	0.0702*	-0.1910***	-0.0267	0.0237	-0.0720*	0.4159***
	Other inactive	0.4159***	1.3209***	0.9747***	0.3284***	0.8961***	0.3883***	0.4165***	0.7215***	0.3009***	0.5436***	0.5030***
Constant		7.4813***	6.1261***	6.8544***	9.2486***	7.5355***	7.3347***	7.6213***	6.4829***	7.7961***	7.6763***	8.4893***
R2		0.2792	0.1240	0.1132	0.2885	0.1367	0.2209	0.3517	0.4446	0.1304	0.3741	0.2257
Observations		9812	19100	5300	4864	6224	12900	5172	7669	7155	5361	7892

Source: EU-SILC 2010, own calculations

5.2.2 Step 2: Shapely value decomposition

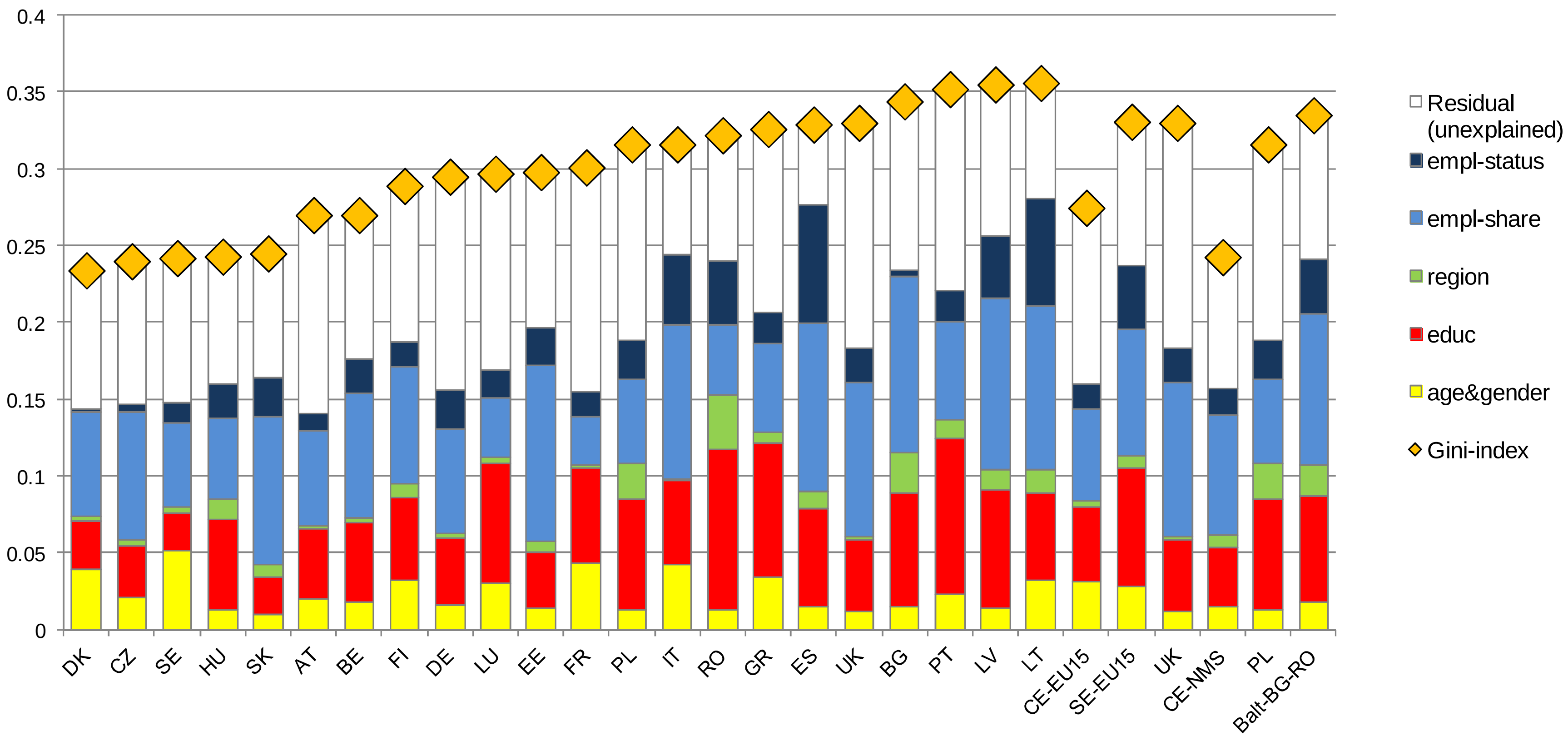
In this section we present the final results for the Shapley value decomposition of income inequality and multidimensional inequality. In addition we apply the Shapley approach also to the Poverty headcount index for the case of household income (see Figure 8).

Figure 6 presents the decomposition results for equivalised household income. On average, about a third of the Gini-index remains unexplained in the analysis. As expected from the simple decomposition analysis in Section 4, the countries where differences between rural and urban regions are an appreciable driver of overall inequality are Poland, Romania, Bulgaria, Latvia and Lithuania and some South European countries. As expected and in line with the results above, educational differences of the heads of households are important drivers of income inequality between households, irrespective of the overall inequality level of the respective country, but much stronger in the Poland, the Baltics, Bulgaria, Romania and South European EU Member States compared to the rest of the EU. The same applies to the individual employment share of households. In all countries analysed this variable contributes strongly positive to income inequality. Especially in crisis-hit countries like Italy, Romania, Spain, Latvia and Lithuania the differences between incomes of households are moreover driven by the employment status of the head of household, simply spoken by the fact that he or she is employed or unemployed/inactive.

The results for the Shapely decomposition of the multidimensional inequality index are presented in Figure 7. For inequality in multidimensional well-being educational differences are the most important single explanatory factor. Only in two countries, Greece and Portugal the age of the head of the household is as important in explaining inequality. Contrary to the results for income inequality, differences between urban and rural areas are not explaining much of total inequality in all countries. The importance of the individual employment shares of the households is somewhat lower in the case of multidimensional inequality compared to inequality of household income. Furthermore, in most of the countries the employment status of the head of the household is not an important predicator for differences in the level of multidimensional inequality.

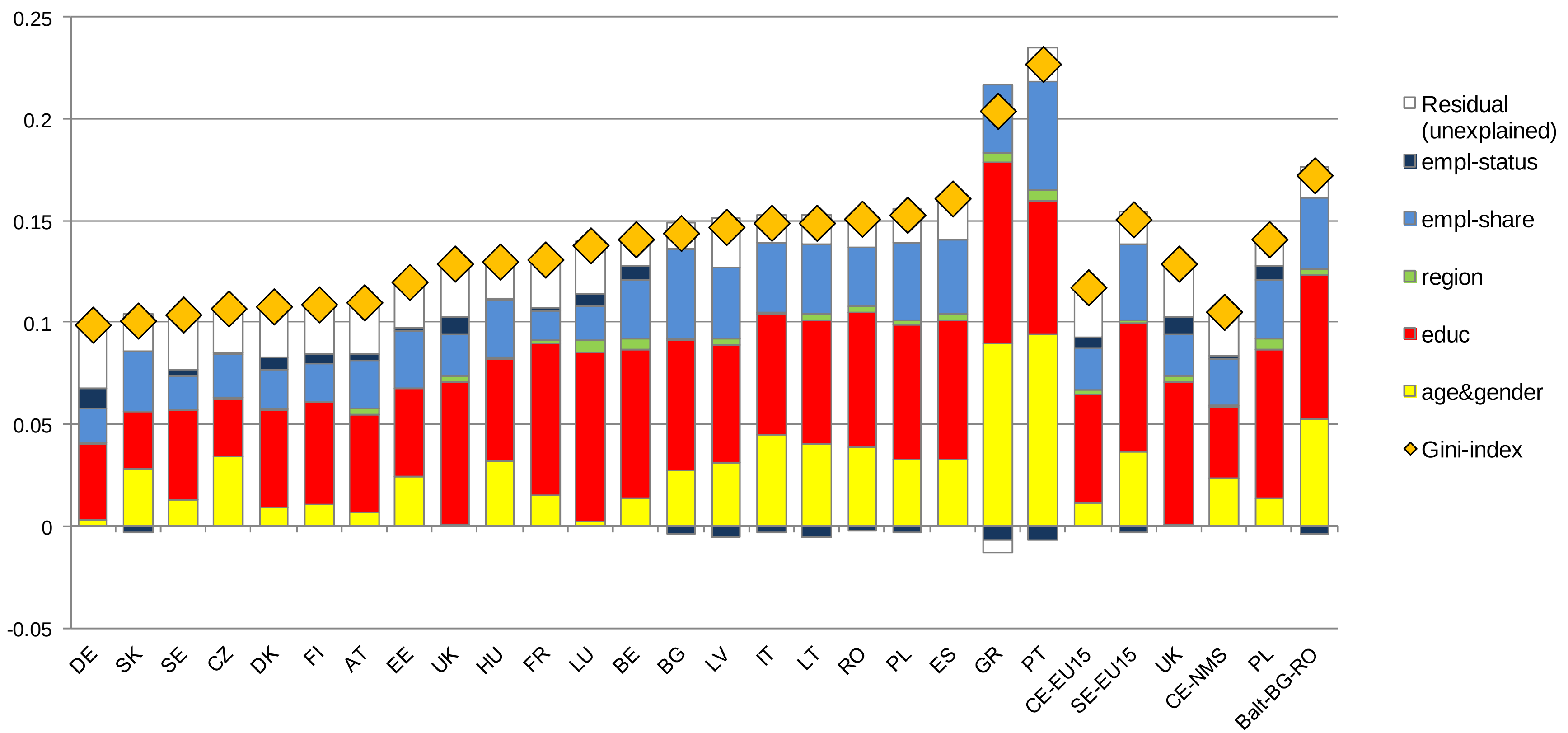
In order to derive a clearer picture of the drivers of inequality we apply the Shapely decomposition approach in the case of income inequality to another, much simpler inequality measure the Poverty headcount index. Households are enumerated as poor if their equivalised per capita income is below 60% of the median income of the total population. In the group of analysed countries the Czech Republic features the lowest poverty rate with about 7% of the households below 60% of the median income, whilst in Bulgaria about 23% of the households can be identified as poor. For the decomposition analysis we can use the same regression results as for the decomposition approach for the Gini index of inequality of household incomes described above. The only difference here is that out of the predicted incomes we do not calculate Gini indices but Poverty indices and thus calculate the marginal contributions of the explanatory variables to those Poverty rates. Figure 8 shows that overall poverty rates can be explained much worse with the applied decomposition approach than in the case of Gini indices. For almost all countries where we have a higher explanatory power (mostly crisis-hit countries) the employment status of the head of the household is the most important explanatory variable.

Figure 6. Shapely value decomposition of inequality in household income
Contributions of variables to total Gini-index



Source: EU-SILC 2010, own calculations

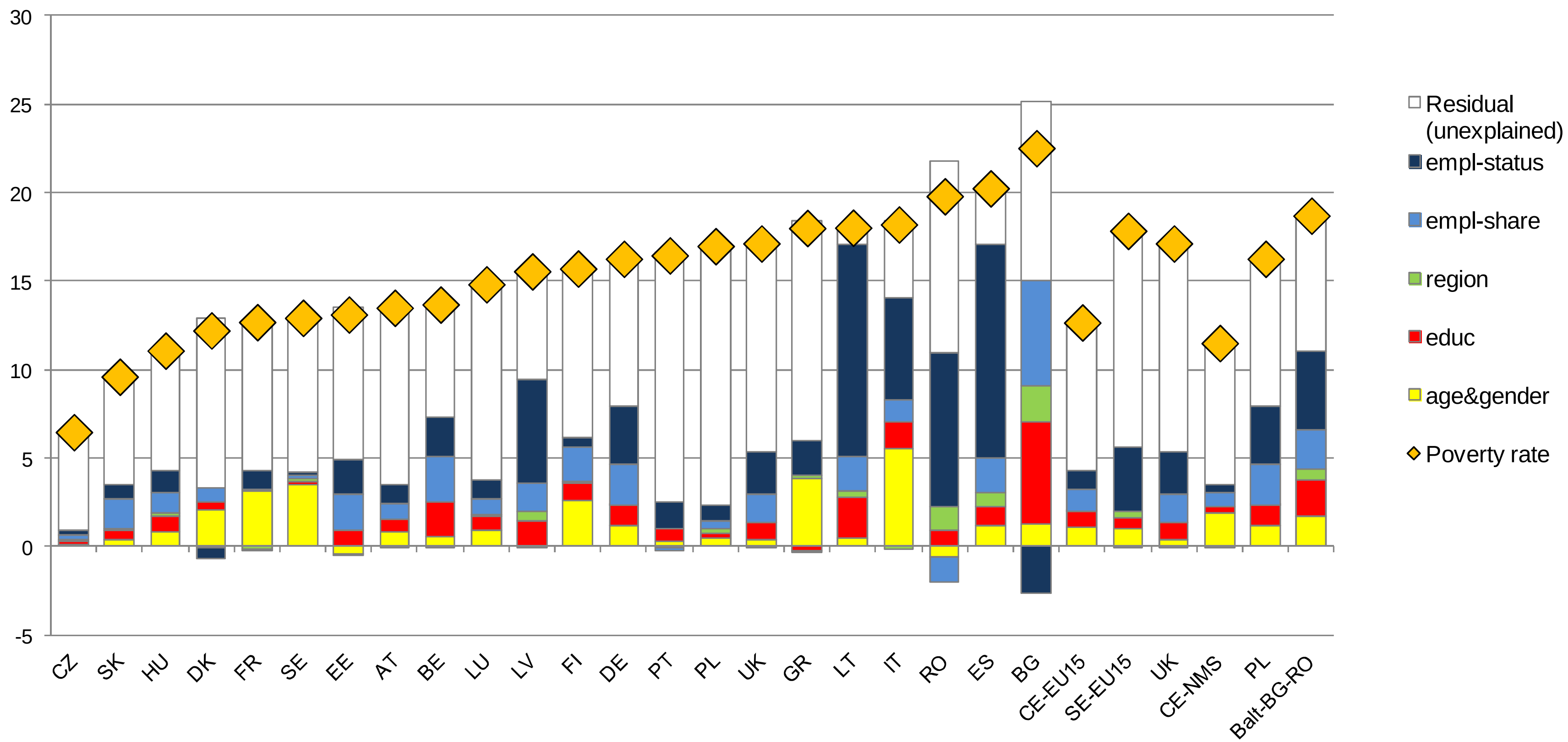
Figure 7. Shapely value decomposition of multidimensional inequality
Contributions of variables to total Gini-index



Source: EU-SILC 2010, own calculations

Figure 8. Shapely value decomposition of poverty in household income

Contributions of variables to total Poverty rate (headcount index: below 60% of median income)



Source: EU-SILC 2010, own calculations

6 Conclusions

6.1 General findings

In this paper we analysed single- and multidimensional inequality in the EU countries for the year 2010. In order to construct a multidimensional inequality index, we included four dimensions: household income, household health, household education level and housing level (quality and space p.c.) and applied various decomposition methods to one- and multidimensional indices of inequality.

Income inequality is, when measured by the Gini index, quite low in most Central European and Scandinavian countries within the EU and highest in East European Member States (Latvia, Lithuania and Bulgaria), South European countries (Portugal, Spain and Greece) and the UK. The inequality in the measured household health status and in the housing indicator did not differ by large between countries; however the CEE EU Member States have above average inequality levels in both cases (except for Slovakia, Slovenia and the Czech Republic in the case of the housing indicator). When it comes to the household educational attainment levels we could see that while inequality is quite low in Central European and Scandinavian countries, but also the Baltic States, the differences are much more pronounced in the South European countries. The combined index of multidimensional inequality highlights the fact, that the South European countries Greece and Portugal feature the highest level of welfare dispersion, while Central European and Scandinavian countries are those with the lowest level of inequality.

In Section 4 we applied standard decomposition techniques to the Mean logarithmic deviation of all four single dimensions and on the multidimensional index as suggested by Massoumi (1986, 1999). The results indicate that household income is most strongly influenced by the education level of the head of household. On average of all countries about 16% of the income differences are accrued from that characteristic, while about 15% of the income differences can be explained by the employment rates of the individual households. Further, also the characteristic employment status of the head of the household explains with about 10% quite a large part of income differentiation. In the case of health inequalities also those households with higher household employment rates and those headed by employed persons faced a somewhat better health status on average. In the case of housing quality the simple decomposition analysis couldn't give much insight into the driving factors of inequality and the results were rather ambiguous across countries. The decomposition results of multidimensional inequality are strongly driven by those attributes with the highest inequality levels, which in our case are the household income and the household education level. On average, differences between education groups account for almost 40% of total differences in household welfare variations. The employment status of the head of the household (on average 17%) and the level of labour market activity of household members (on average 21%) also exert a strong influence on the level of well-being, less so the age of the head of household (13%). The other characteristics of the households analysed have only minor effects.

In Section 5 we applied the Shapely value decomposition to household income inequality, the multidimensional inequality measure and a poverty index of household income inequality. This method is based on a regression approach which allows considering all explanatory variables simultaneously and conditional on each other. Further the Shapley value approach allows calculating the contribution of groups of these variables to the respective inequality measure. This approach seems to work best for the Gini coefficient. In the case of income inequality we can explain on average about two third of the Gini index by our approach. In general the household specific employment rate is the most important driver of income inequality in the EU countries. Second most influential are differences in the educational attainment rate of the head of the household. In all countries the combined effect of gender and age is explaining just a small part of the inequality levels. The same is the case for differences between urban and rural areas for most of the EU countries. Only in Poland, Bulgaria and Romania regional differences are remarkable additional drivers of the level of income inequality.

The results of the decomposition approach look somewhat different, when we apply the method not to inequality of income but to the multidimensional case. In many countries the importance of education as a trigger of inequality increases, while the impact of differences in labour market participation declines. Also differences between rural and urban areas (also for those countries, where important for income inequality) become negligent. In some countries, especially Portugal and Greece, age differences are strongly driving overall inequalities of multidimensional welfare.

We not only applied the decomposition approach to inequality indices that consider the whole income distribution, but also to a Poverty index, which takes into account only a distinct part of the distribution (in our case we used a headcount index for all those households with an equivalised per capita income of less than 60% of the median income). However, the results were somewhat disappointing since the decomposition analysis could for most countries only explain a very small part of the existing poverty levels. For those countries where we have a higher explanatory power (mostly countries that are hit hard by the economic crisis) the employment status of the head of the household is the most important variable driving overall income poverty.

6.2 Inequality and poverty in the CEE EU Member States in comparison

The comparative analysis of single- and multidimensional inequality and poverty in respect to the CEE EU Member States shows the following:

- Regarding income inequality the CEE EU Members comprise different subgroups, the first consisting of the Czech Republic, Slovenia, Slovakia and Hungary, which very much resemble features of Scandinavian/Central European - EU15 countries. Relatively low income differences between households are mostly driven by disparities of labour market participation. However, in those Central European New EU Member States differences between rural and urban regions are an additional driver of income inequality. The highest levels of income inequality in the EU are to be found in Latvia, Lithuania and Bulgaria while Estonia, Poland and Romania also have levels above the EU 27 average. This group of New EU Member States resembles features comparable to South European countries. Their higher levels of income inequality are apart from differences in labour market participation driven in addition by variations between educational attainment groups. Furthermore households at the country side have on average lower income levels compared to those in urban areas (also conditional on all other factors accounted for as shown by the Shapley value decomposition analysis). The analysis of poverty levels and their decomposition did not deliver additional insights.
- Regarding further attributes of our multidimensional inequality analysis the results are as follows. Inequality according to our constructed conditional health status of households is in general lower compared to income differences. However, all CEE EU Member States have above EU-average levels of inequality comparable only with those of Greece and Portugal. The decomposition analysis did not deliver much insight into the sources of inequality although the health status in the CEE and South European countries are influenced more strongly by variables describing labour market participation.

- Inequality according to our housing indicator (including information on housing quality and space p.c.) is also at a much lower level compared to income inequality. Again the CEE EU Member States comprise of two different subgroups the first including Slovakia, Slovenia and the Czech Republic with a very low level of housing inequality and a second subgroup containing the Baltic States, Bulgaria, Romania, Poland and in this respect also Hungary. The decomposition approach did not tell us much about the sources of inequality in this attribute of the multidimensional analysis in the CEE region, only in some countries like Bulgaria and Romania stronger differences between educational attainment groups can be detected.
- Educational differences between households are generally much lower in the CEE EU Member States compared to South European countries. However also in the group of CEE countries the situation is disperse. Especially in Romania and Poland (to a lesser extent in Bulgaria and Lithuania) a substantial share of the population can be characterized as poor according to their educational attainment levels (below 60% of the median educational level of households). In general in CEE countries educational segregation is as eminent as in the rest of the EU. However, stronger educational differences are also evident between urban and rural households in Bulgaria, Romania but also Hungary and Slovakia.
- The analysis of multidimensional inequality sums up the findings of the various attributes discussed above. Thus, again Slovakia and the Czech Republic feature very low inequality levels comparable to Scandinavian/Central European - EU15 countries closely followed by Estonia, Slovenia and Hungary. Bulgaria, Latvia, Lithuania, Romania and Poland form a subgroup of countries similar to South Europe. However, Greece and Portugal feature inequality levels that are much above those of the latter subgroup of CEE EU Member States. Particularly differences between age and educational attainment groups are more pronounced in this latter subgroup of countries.
- The analysis shows that with respect to income and multidimensional inequality the region of CEE EU Member States comprises of at least two distinct country groups. The first consists of the Czech Republic, Slovakia and Slovenia, which feature low levels of inequality in all attributes (except for our constructed indicator of conditional health status) when compared with the rest of the EU. The second group, comprising Bulgaria, Romania, Poland and the two Baltic countries Latvia and Lithuania, has according to all attributes (except for educational attainment levels) inequality levels at the upper end of the ranking of EU countries. The two countries in-between are Hungary and Estonia, the first featuring low levels of income inequality, but quite high levels of inequality in the attributes health and housing. Estonia, although having a high level of inequality in respect to the housing indicator and a level of income inequality resembling the EU average, however features a low level of inequality according to educational attainment levels of households.

References

Atkinson, A. B. (1975). The Economics of Inequality. Oxford: Clarendon Press.

Atkinson, A. B. and E. Marlier (eds.) (2010). Income and living conditions in Europe. Luxembourg: Publication Office of the European Union.

Bourguignon, F. (1999). Comment to: Multidimensional approaches to welfare analysis. In J. Silber (Ed.), Handbook of Inequality Measurement. Kluwer, M.A.

Cowell, F. A. and C. V. Fiorio (2009). Inequality decompositions. A reconciliation. ECINEQ Working Paper Series WP 2009-117.

Israeli, O. (2007). A Shapely-based decomposition of the R-square of a linear regression. Journal of Economic Inequality, 199-212.

Jenkins, S. P. (1995). Accounting for inequality trends: Decomposition analysis for the UK, 1971-86. Economica 62, 29–63.

Justino, P. (2005). Empirical Applications of Multidimensional Inequality Analysis, PRUS Working Paper, No. 23.

Leitner, S. and R. Stehrer (2009). Determinants of inequality in selected SEE countries: Results from Shapley value decompositions. GDN research paper, mimeo.

Lugo, M. A. (2005). Comparing multidimensional indices of inequality: Methods and applications. ECINEQ Working Paper 2005-14.

Maasoumi, E. (1986). The measurement and decomposition of multidimensional inequality. Econometrica, 991-997.

Maasoumi, E. (1999). Multidimensional approaches to welfare analysis. In J. Silber (Ed.), Handbook of Income Inequality Measurement. Boston: Kluwer Academic.

Molini, V. and G. Wan (2008). Discovering sources of inequality in transition economies: a case study of rural Vietnam, Journal of Economic Change, 41, 75-96.

Savaglio, E. (2006a). On multidimensional inequality with variable population size, Economic Theory 28, 85-94.

Savaglio, E. (2006b) .Three approaches to the analysis of multidimensional inequality, in Farina, F. and E. Savaglio (eds), Inequality and Economic Integration, Routledge, London.

Sen, A. (1997). On Economic Inequality. Oxford, UK: Oxford University Press.

Shorrocks, A. F. (1999). Decomposition procedures for distributional analysis: A unified framework based on the shapely value. Draft paper, University of Essex.

Wan, G. (2002). Regression-based Inequality Decomposition: Pitfalls and a Solution Procedure. UNU-WIDER, Discussion Paper No. 2002/101.

Wan, G. (2004). Accounting for income inequality in rural China: A regression-based approach. Journal of Comparative Economics 32, 348–363.

Weymark, J. A. (2004). The normative approach to the measurement of multidimensional inequality. Working Paper 03-W14R, Vanderbilt University, Nashville.

Appendix Tables

Table A.1. Regression results: Dependent variable: Logarithm of Multidimensional welfare indicator (modified OECD equivalence scale)

		AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GR
Age		0.0036**	0.0096***	0.0094***	0.0069***	-0.0010	0.0122***	-0.0022	0.0151***	0.0083***	0.0117***	0.0394***
$Age^2/100$		-0.0044***	-0.0105***	-0.0129***	-0.0113***	0.0001	-0.0124***	-0.0014	-0.0181***	-0.0097***	-0.0126***	-0.0446***
Male		-0.0041	0.0539***	0.0489***	0.0263***	-0.0020	0.0129*	0.0452***	0.0174**	0.0116**	0.0274***	0.1771***
Education level	Upper secondary	0.2286***	0.2728***	0.2739***	0.1828***	0.1633***	0.1818***	0.1868***	0.2263***	0.1391***	0.2452***	0.3558***
	Tertiary	0.3642***	0.4059***	0.4172***	0.3349***	0.2998***	0.3111***	0.3280***	0.3482***	0.2970***	0.4176***	0.5007***
Household employment share		0.1362***	0.2110***	0.3378***	0.1499***	0.1097***	0.1547***	0.2000***	0.2673***	0.1363***	0.1224***	0.2962***
Rural region		0.0412***	0.0553***	-0.0057	0.0176***	0.0171***	0.0241***	0.0074	-0.0280***	0.0116*	0.0232***	-0.0399***
Employment status	Employee	0.0188	-0.1370***	-0.0705***	-0.0700***	0.0247**	-0.0350	-0.0644***	-0.0287*	-0.0077	-0.0510***	0.0744**
	Self-employed	0.0701***	0.0143	-0.0441**	0.0297*	0.1354***	0.0794***	-0.0047	-0.0123	0.0496***	0.0249*	-0.0966***
	Retired	0.0701***	-0.0128	-0.0206	0.0491***	0.1680***	0.0684**	-0.0017	-0.0475***	0.0621***	0.0532***	-0.0833**
	Other inactive	0.0704***	0.0834***	0.1255***	0.1108***	0.1513***	0.1462***	0.0167	0.0682***	0.1003***	0.0908***	0.0756**
Constant		-1.4305***	-1.6311***	-1.8434***	-1.5278***	-1.3425***	-1.5875***	-1.3808***	-1.6584***	-1.4178***	-1.6283***	-2.3841***
R2		0.3333	0.4230	0.3979	0.3532	0.3619	0.2948	0.4073	0.4118	0.3632	0.4427	0.4970
Observations		6104	6048	6152	9065	12700	5618	4937	13400	10300	10900	6800

		HU	IT	LT	LU	LV	PL	PT	RO	SE	SK	UK
Age		0.0084***	0.0251***	0.0117***	0.0004	0.0030	0.0124***	0.0332***	0.0170***	0.0131***	0.0082***	0.0022
$Age^2/100$		-0.0129***	-0.0275***	-0.0167***	0.0008	-0.0077***	-0.0165***	-0.0408***	-0.0200***	-0.0133***	-0.0122***	-0.0017
Male		0.0289***	0.0562***	0.0644***	-0.0072	0.0420***	0.0570***	0.1447***	0.0490***	0.0319***	0.0316***	0.0031
Education level	Upper secondary	0.2106***	0.2391***	0.2714***	0.2611***	0.2424***	0.3306***	0.3779***	0.2565***	0.2062***	0.1661***	0.2418***
	Tertiary	0.3707***	0.3773***	0.4288***	0.4346***	0.4111***	0.5032***	0.5936***	0.4455***	0.3290***	0.2973***	0.3748***
Household employment share		0.2072***	0.2521***	0.2564***	0.1547***	0.2697***	0.3032***	0.4233***	0.2446***	0.1344***	0.2208***	0.1363***
Rural region		-0.0151**	-0.0104*	-0.0271**	0.0546***	-0.0297***	-0.0203***	-0.0532***	-0.0280***	0.0023	0.0000	0.0265***
Employment status	Employee	-0.0484***	0.0543***	-0.0864***	0.0483**	-0.0406*	0.0339*	-0.0706	-0.1022**	-0.0108	-0.0517**	-0.0314
	Self-employed	-0.0003	0.0040	-0.0369*	0.0929***	-0.0283	-0.0071	-0.1059***	0.0009	0.0465**	-0.0016	0.0770***
	Retired	0.0287	0.0262	-0.0308	0.0811***	-0.0507*	-0.0234	-0.1013**	-0.0556*	0.0679*	-0.0269	0.0967***
	Other inactive	0.0863***	0.1358***	0.0970***	0.1242***	0.0767***	0.1142***	0.0809*	0.1247***	0.0955***	0.0910***	0.1002***
Constant		-1.6537***	-1.9744***	-1.8020***	-1.4166***	-1.6194***	-1.9420***	-2.2705***	-2.0358***	-1.6160***	-1.5709***	-1.4583***
R2		0.4182	0.4128	0.4114	0.5045	0.3596	0.4634	0.4137	0.4474	0.3346	0.3989	0.3456
Observations		9752	18900	5290	4855	6179	12900	5169	7646	6949	5208	7854

Source: EU-SILC 2010, own calculations