Working Paper Series

Ranking populations in terms of Inequality of health opportunity: A flexible latent type approach

Paolo Brunori
Caterina Francesca Guidi
Alain Trannoy

ECINEQ WP 2020 - 515

ECINEQ 2020 - 515
January 2020

www.ecineq.org

Ranking populations in terms of Inequality of health opportunity: A flexible latent type approach [‡]

Paolo Brunori
University of Florence & University of Bari

Caterina Francesca Guidi
European University Institute

Alain Trannoy
Aix-Marseille University (Aix-Marseille School of Economics), EHESS and CNRS

Abstract

We offer a flexible latent type approach to rank populations according to unequal health opportunities. Building upon the latent-class method proposed by Li Donni et al. (2015), our contribution is to let the number of types vary to obtain an opportunity-inequality curve for a population that gives how the between-type inequality varies with the number of types. A population A is said to have less inequality of opportunity than population B if its curve is statistically below that of population B. This version of the latent class approach allows for a robust ranking of 31 European countries regarding inequality of opportunity in health.

Keywords: Inequality of opportunity, health inequality, latent class, opportunity-inequality curve, self-assessed health.

JEL Classification: I14, D63.

[‡]We would like to thank Raymond Schmelzer for valuable comments.

1 Introduction

The scope of equality of opportunity is different from the scope of equality of outcomes by the very fact that one is interested in the process generating the diversity of results. According to this view, valuable outcomes should not be influenced by circumstances beyond individual control or preference. This idea has been framed in various versions (Dworkin, 1981; Arneson, 1988; Cohen, 1989) and has led to distinct definitions of equality of opportunity in the economic literature (Roemer, 1998 and Fleurbaey 2008). They concern a variety of welfare dimensions or "outcomes," such as income, wealth, consumption, education, and health. The interested reader can refer to some recent literature reviews (Roemer and Trannoy (2015), Ramos and Van de Gaer (2016)).

In very recent years a number of empirical studies have followed the inequality of opportunity approach to quantify unfair inequality in health (Rosa Dias 2009 and 2010; Trannoy et al. (2010); Garcia-Gomez et al. (2012), Van De Gaer et al. (2012); Jusot et al. (2013), Bricard et al. (2013); Li Donni et al. (2014); Carrieri and Jones (2016)). These studies represent a link between the literature on socioeconomic inequalities in health and the social choice literature on equity, responsibility and compensation. The former has investigated the role of income, wealth and education on the disparity of health status (see for instance, Lantz et al., 2001; Burstrom et al., 2005; Kunst et al. 2005, Mackenbach et al., 2008; Marmot et al., 2012; Toch-Marquardt et al., 2014, Hu et al. 2017).

This paper represents a contribution to this growing literature. We estimate inequality of opportunity in health in a large sample of countries (31 European countries). Our estimates are based on the 2011 round of the European Union Survey on Income and Living Conditions (EU-SILC) and are obtained by adopting a new estimation method. We extend the latent class approach proposed by Li Donni et al. (2015) and Carrieri et al. (2019) by introducing a sequential dominance criterion to rank countries regarding the inequality of opportunity in health.

The Roemer concept of type (Roemer 1998) is central to the equality of opportunity analysis. Belonging to a type means that people share the same determinism regarding the outcome variable under consideration. The type concept is key to measure inequality of opportunity because it helps to link inequality of opportunity measurement and measuring inequality between subgroups. In the most common sense, inequality of opportunity is measured as the between-type inequality. To this end, a counterfactual society is built in which within-type inequality is neutralized.

The most immediate way of defining type is to consider that they are elements of the Cartesian product of all subsets of observable variables describing circumstances beyond individual control. For instance, if there are ten ethnic groups and two gender groups, then there will be 20 types. There are two problems of doing that way, the rich data set problem and the unobservable characteristics. On the one hand, with a rich data set, the procedure leads to too many types. On the other hand, the procedure does not count for unobservable characteristics making the within-type heterogeneity too large. Both issues are not presumably pregnant in the same way for every data set. A small data set will likely suffer more from the later problem.

The latent type approach provides a statistical procedure that deals with both issues. The key idea is the following. If types provide a sound partition of the population, it must be true that the within-type inequality of circumstances should be minimal because people obey the law of the same social determinism. Conversely, the between-type inequality of circumstances should be maximum in some sense.

In the latent type approach, observable circumstances are considered as “manifest variables" of the membership to latent social types. Let the number of types exogenously given. Individuals are assigned to *a* type so that, in each group, the correlation of observable circumstances is the lowest possible. In technical terms, individuals are associated with each type to maximize local independence. Perfect local independence means that conditional on type membership, all of the observable circumstances are statistically independent. Therefore, latent types are identified so to minimize within-type homogeneity in terms of circumstances, that is, maximizing between-type heterogeneity in terms of circumstances.

The latent class analysis does not indicate the number of latent classes one should use. Li Donni et al. (2015) suggest selecting the most appropriate number of types by adopting a parsimony criterion able to balance between over- and under-fitting the model. We instead propose a flexible method regarding the number of types. Our methodological contribution is here to let the number of types vary to obtain *an opportunity-inequality curve* for a population that gives the between-type inequality in a range of type numbers. A population A is said to have less inequality of opportunity than population B if its curve is statistically below that of population B. This version of the latent type approach allows for a robust ranking of countries regarding inequality of opportunity.

We apply this methodology to rank 31 European countries according to a measure of self-assessed health status (SAH). We focus on health conditions when people are active on the job market. SAH is widely used and has shown to be a reliable indicator of general health and well-being (Jylha 2009). Even if the level of this indicator is not related to mortality rates at the country level, it is a reliable indicator of general health and well-being at the individual level (Idler and Benyamini 1997; Huismans et al. 2007; Bopp et al. 2012). We find a Western/Eastern divide where Eastern Europe countries are bottom ranked. Mediterranean countries such as Spain, Malta, Greece and Cyprus are among the best performers. The method is also useful to investigate the reliability of the indicator of subjective health in EU-SILC and some strange patterns seem to occur for a few countries.

The closest research was those of Bricard et al. (2013), who also aim at rank European countries according to SAH. They mainly use the third wave of the Survey of Health, Ageing and Retirement in Europe (SHARE) which was collected in 2008/2009, called the Retrospective Survey of SHARELIFE– as it focuses on people’s life histories of people aged 50 and over. The variance of SAH is decomposed into four components: due to demographic characteristics, due to effort/choices, due to circumstances, and an unexplained residual for a subset of 13 European countries. *IOpH* is the share of variance systematically correlated with circumstances. Inequality of opportunity in health is present in every surveyed country. However, they do not find a clear geographical pattern as we find with Eastern European countries less favorably ranked than Mediterranean countries. The limited set of countries and the focus on a different age group may explain this different pattern, without totally excluding that it may also partly come from the difference in the empirical methodologies.

The remaining part of the paper is structured as follows: Section 2 is the methodologic section. It introduces the model used to define and measure equality of opportunity in health, *IOpH,* and presents the latent-class estimation of types. It then introduces our main methodological innovation, the opportunity-inequality curve, to compare inequality of opportunity across populations. Section 3 presents the data set, our estimates of the ranking of countries. Section 4 discusses the findings and their reliability. Section 5 concludes.

2 Model and type estimation

2.1 A reduced-form model

We introduce the canonical equality of opportunity model. Consider a population of N individuals over which distribution of health conditions H is defined. We assume that individual health, h is determined by three types of traits: a finite set of factors over which individuals have control (E) which are called ``effort'' variables, a set of (avoidable) factors for which individuals cannot be held responsible ($C$), which are called ``circumstances'', and a set of unavoidable determinants of health, for example, age (D).

The individual outcome is generated by a function of circumstances and responsibility variables:

$$h = g(C, E, D) \tag{1}$$

All the possible combinations of circumstances' values taken one at a time from C define a partition of the population into types. Individuals belonging to the same type are characterized by identical circumstances.

In this study, we do not introduce effort (lifestyle) in the model, meaning that effort variables are among the unobservable. It means that the fraction of effort correlated to circumstances or demographic variables would be added somewhat to the impact of these variables on the health status. The econometric strategy is first to clean the effect of demographic variables and actually age to obtain a relationship between health and circumstances. Inequality of opportunity is then evaluated in the counterfactual distribution obtained by replacing the health variable at the individual level by the type's average. Hence, the inequality of opportunity in health is simply defined by between-type inequality.

Formally, if $\widetilde{H}$ denotes the counterfactual distribution, we have

$$IOpH = I(\widetilde{H}) \tag{2}$$

where *I()* is a suitable absolute inequality measure. Consistently with other contributions that deal with binary outcomes in the literature, we adopt an inequality measure based on absolute outcome differences (Trannoy et al., 2010; Bricard et al., 2013). We estimate IOpH as the dissimilarity index of the counterfactual distributions proposed by Paes de Barros et al. (2009).

$$D = \frac{1}{2\mu} \sum_{l=1}^{L} w_l \left| \mu - \tilde{h}_l \right|, \tag{3}$$

where w_l is the share of group *l* in total population and *L* is the number of types.

In order to implement cross-country comparisons in terms of *IOpH*, we need to identify Romerian types, namely groups of individuals sharing the same circumstances.

Recently, Li Donni et al. (2015) have proposed to determine the partition in types based on latent class model. The underlying assumption is that observable circumstances are only a subset of the full set of relevant circumstances, some being non-observable. To take into consideration both types of circumstances, they consider observable circumstances as "manifest variables" of a latent categorical variable: the membership to latent types. Types membership is then determined by estimating a latent class model. Given a number of types exogenously determined, individuals are associated with each type so to maximize local independence. Perfect local independence means that conditional on type membership, all of the manifest variables are statistically independent.

2.2 Local independence through an example

Type membership is estimated by maximum likelihood under the assumption of local independence. Here we provide an intuition of what this assumption means; For a formal explanation see the Online appendix [link].

The following table gives the contingency table of the distribution of pupils according to the education of the parents.

Table 1a: Contingency table for the whole population

Father Edu/Mother Edu	Low	High	Total
Low	414	10	**424**
High	422	154	**576**
Total	**836**	**164**	**1000**

To understand the local independence assumption, we start with the trivial case where the number of latent types is precisely equal to the number of all four possible combinations of categories. Of course, in that case, it is very natural to identify the latent-type partition to the potential type partition. Namely, the posterior probability of belonging to a latent type for a given association of categories is 1 for one and only one latent type. We deduce that for type 1, corresponding to the low/low association, we get the following allocation of individuals.

Table 1b: Contingency table for Type 1 with 4 latent classes

Father Edu/Mother Edu	Low	High	Total
Low	414	0	414
High	0	0	0
Total	**414**	**0**	**414**

Let us remark that the probability of observing any combination of parental education is just the product of the marginal probabilities of observing any level of parental education. For instance, the probability of observing the education level high/low is 0. It is the product of the probability of observing high father education which is zero and the probability of observing low father education, which is also zero.
The same property holds for any latent type in this trivial case. In each latent type, the conditional probability of observing an individual with a particular association of any category is given by the product of the marginal probabilities. This property is called *local independence*.
Now suppose that we want to reduce the number of latent types to only two. Clearly, these latent types will necessarily mix people with different categories. It cannot be otherwise. How to mix the people in each latent type? Well, an idea is probably to keep the local independence property that holds in the trivial case. It means that for an individual of a given type, the probability of having a

mother with a low education does not depend on the probability of having drawn a high education father or a low education father. Knowing the father's education does not allow us to predict mother education. The probability of having a high education mother is exactly given by this probability in that particular type. In that sense, the individuals in this latent class remain similar. They are only characterized by the type probability of having a high or a low education parent. Ex post, individuals belonging to a given latent type will be different but ex-ante they have precisely the same probability of being in each cell.

Let us illustrate how this local independence assumption allows us to define the two-latent types. The latent type estimation procedure is producing the probability for each cell (low-low, high-low, etc) of belonging to latent type.

Table 1c Joint probabilities

LATENT CLASS 1:

Father Edu/Mother Edu	Low	High
Low	**0.0154**	**0.5608**
High	**0.4076**	**0.9825**

LATENT CLASS 2

Father Edu/Mother Edu	Low	High
Low	**0.9846**	**0.4392**
High	**0.5923**	**0.0175**

Applying these probabilities to the numbers of Table 1, we get the following allocation of individuals for the two types.

Table 1d. Contingency table for 2 latent classes

LATENT CLASS 1

Father Edu/Mother Edu	Low	High	Total
Low	6	6	12
High	172	151	323
Total	**178**	**157**	**335**

LATENT CLASS 2

	Low	High	Total
Low	408	4	412
High	250	3	253
Total	**658**	**7**	**665**

The reader can check that local independence is satisfied for each cell for both latent types. For instance, the probability of drawing low/low in latent type 2 = 408/665= 0.61 is equal to the probability of the second-type probability of drawing a low education father 658/665=0.99 and the second-class probability of drawing a low education mother 412/665 =0.62

2.3 The opportunity-inequality curve

Once the allocation of individuals in L latent types is done, we are able to predict the counterfactual distribution as the distribution of types' mean outcome: $\widetilde{H}_{,L}$. *IOpH* can, therefore, be estimated for any possible number of latent types. How to optimize the choice of L?

To address this issue, Li Donni et al. (2015) suggest selecting the number of latent types guided by an information criterion such as the Bayesian criterion (BIC). Such a criterion evaluates the likelihood of the model introducing a penalty term for the number of parameters estimated. The BIC selects the most appropriate model balancing between choosing a model able to closely fit the data in the sample, and choosing a model with the lowest possible sampling variance.

We here suggest that BIC may not be the best criterion to select the L when willing to estimate inequality of opportunity. Besides, to select an exact number of latent types may not be necessary. We develop the two ideas below.

For a latent class model, a perfect fit is obtained when the distribution of manifest variables is orthogonal to classes, namely when local independence is fully satisfied. The BIC of a latent class model, therefore, evaluates the model's ability to explain the correlation of manifest variables in the sample. However, when estimating inequality of opportunity, the aim is not to explain covariance of circumstances, but to identify the partition in types that most explain the outcome variability. It is a specific case of the more general problem of using latent class membership as a predictor for a distal dependent variable. As discussed by Lanza et al. (2013) such an approach is likely to produce attenuated estimates of the effect of the manifest variables on the outcome, here, a downward biased estimate of inequality of opportunity.[2]

Therefore it seems more appropriate to select the number of latent types considering the ability of the resulting partitions in types to explain the outcome variability. It can be done by *k*-fold cross-validation to directly obtain a nearly unbiased measure of the true out-of-sample prediction error of alternative models (James et al., 2014). The preferred number of types is the number that minimizes the expected squared error in predicting individual outcome as function of circumstances. Such a procedure balances between the need to maximize the outcome variability explained by types membership (avoiding downward biased IOpH estimates) and the risk to overfit the model choosing an unreasonably high number of types (that would result in upward biased estimates).

[2] This is shown to be the case in a number of samples by Brunori et al. (2018) that use the approach suggested by Li Donni et al. (2015) to measure IOp in income.

More precisely, to estimate the expected mean-squared error, MSE, by k-fold cross-validation we randomly partition the sample in k equal sized subsamples. A single subsample is removed from the sample (test sample) and the remaining *k-1* subsamples (training sample) are used to estimate the latent type model. Based on the parameters obtained in the training sample, the latent type membership is predicted for individuals in the test sample. We then regress the health status on the list of fixed effects, once for each latent type, in the test sample. The mean square error is then $MSE = \frac{1}{L}\sum_{l=1}^{L}\frac{1}{n_l}\sum_{i=1}^{n_l}(\mu^l - h_i^l)^2$ where n_l and μ^l are the number of individuals and the mean outcome of latent type l respectively. This process is repeated k times, using each one of the k subsamples as the test set. The average of the k MSEs is then used to compare models' predictive performance. The number of latent types that produces the lowest MSE, L^*, is our preferred number of latent types.

Note, however, that it is not necessary to select an exact number of latent types when estimating inequality of opportunity. If one is unsure about the real number of latent types, as it is inevitably in the case in practice, it may be of interest to check how inequality of opportunity evolves when a different number of latent types are considered. Such a sequential estimation procedure is particularly useful when the interest is to rank different populations for which criteria goodness of fit and parsimony may suggest letting the latent type number vary. The sequential comparison of inequality indexes provides a criterion to rank population which is robust to the number of latent types assumed.

The sequential procedure consists of the following steps:

- *i)* set *L=1* (this is a trivial case because all individuals belong to the same latent type and by construction inequality of opportunity is zero);
- *ii)* estimate a model with L types;
- *iii)* regress individual outcome on latent-type membership indicators;
- *iv)* assess the out of sample prediction error, MSE_L, of the model estimated in *ii)* by k-fold cross-validation;
- *v)* increase L by one unit;
- *vi)* if *L=2* repeat steps *i – iv;*
- *vii)* if *L≠2* and $MSE_{L-1} \geq MSE_L$: repeat steps *i – iv;*
- *viii)* if $MSE_{L-1} < MSE_L$: stop.

At the end of the sequential procedure, *L-1=L** is the number of latent classes that minimizes the expected mean square error out-of-sample. This number will typically differ across countries. Among all optimal L^*, we indicate the highest as L^{MAX}. We then estimate *IOpH* as between-type inequality for all countries and all number of latent types from one to L^{MAX}.

Not knowing the real number of latent types, when comparing different populations (e.g. *A, B*), in terms of *IOpH,* we rank A as more unequal in terms of health opportunity then B when, for all $L \leq L^{MAX}$, inequality of opportunity is larger (or equal) in A than in B and it is strictly larger for at least one value of L:

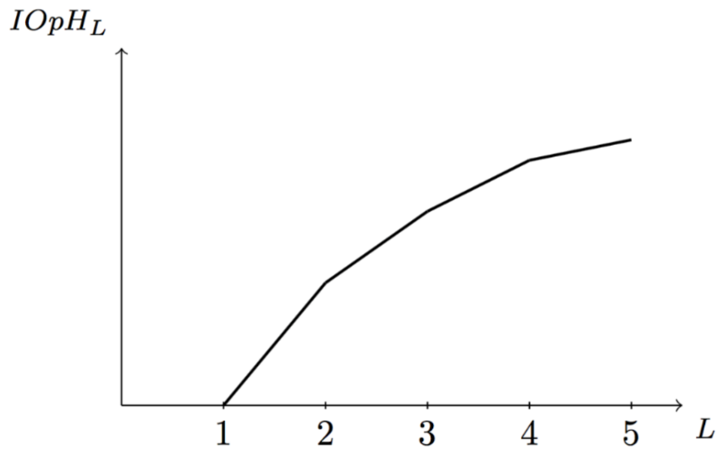
$$IOpH^A > IOpH^B \iff \forall\, L = 1, \dots, L^{MAX}\;\; IOpH_L^A \geq IOpH_L^B \text{ and } \exists L\; s.t. IOpH_L^A > IOpH_L^B \qquad (5)$$

The ranking criterion describes a dominance procedure, which involves comparisons of *IOpH* in the two samples for every possible number of latent types $L \leq L^{MAX}$.

The intuition of how this method ranks populations can be clarified considering the curve plotted in Figure 1 the *opportunity-inequality curve* hereafter. On the horizontal axis, we draw the number of latent classes considered, in the example: *L=1,...,5*. On the vertical axis, we report $IOpH_L$, the inequality in the counterfactual distribution obtained assuming L latent classes.

The first coordinate of the opportunity curve is always (1,0). When only one latent type is assumed, between-type inequality, $IOpH_1$, is zero. Then the inequality of opportunity curve will generally (but not necessarily) increase in with *L*.

Figure 1: Opportunity-inequality curve for latent types from 1 to 5



The ranking criterion in Eq. (5) can then be defined in terms of opportunity curve: distribution A has higher *IOpH* than distribution *B* if the opportunity curve of *A* never lies below the opportunity curve of distribution *B* and lies above for at least one number of latent types.

This dominance criterion gives a partial ordering that makes the ranking relation in terms of *IOpH* incomplete. It is nevertheless possible to obtain a complete ranking of distributions remaining agnostic about the real number of latent types in the population if we define *IOpH* as the average of all measures obtained with all possible number of classes:

$$IOpH = \sum_{L=1}^{L^{MAX}} w_l IOpH_L \tag{8}$$

Note that geometrically *IOpH* is equal to the Riemann Sum of L^{MAX} rectangles divided by L^{MAX} i. e. an approximation from below of the exact area under the opportunity-inequality curve divided by L^{MAX}. Note also that in empirical implementations, the uncertainty of the estimated *IOpH* will increase with the number of latent types. Therefore, one may prefer to use the weighted average of the vector of inequality measure, using weights, for instance, decreasing with the number of types.
In what follows, we will estimate $IOpH_L$ for each meaningful number of latent types. We will then check whether we can rank countries independently from the number of types considered, testing the dominance criterion in Eq. (5). Finally, we will obtain a complete ranking of countries comparing the area below the inequality of opportunity curve of each country.

3. Data and Results

We need data about the distribution of health conditions in the population that are representative of the entire population of as many European countries as possible and contain information about health-relevant individual circumstances. Such requirements restrict the possible choices. We opt for the 2011 wave of the European Union Statistics on Income and Living Conditions (EU-SILC). The EU-SILC is an annual survey, coordinated by Eurostat to produce harmonized statistics on income and living conditions in Europe. The survey also contains a few questions about the respondents' health outcomes. Together with this information, EU-SILC 2011 contains an ad-hoc module about the intergenerational transmission of disadvantages. This special module aims to allow researchers to study aspects of inequalities persistence across generations. To this end, the module contains variables describing the socioeconomic background of all respondents between 25 and 59 years of age.

3.1 Data

EU-SILC, released in 2011, contains information on a few health aspects for 31 European countries. Three questions deliver an indicator of health status: self-assessed general health, self-reported chronic illness, limitation in activities because of health problems. Consistently with previous studies, including, Kunst et al., (2005), Rosa Dias, (2009); Li Donni, (2014); Lazar, (2013); Bricard et al., (2013), we select the first as our preferred measure of health outcome. The question asked, "How is your health in general?" and have six possible answers: "very good", "good", "fair", "bad", "very bad", "don't know".[3]
We dichotomize the answers by collapsing those that reported at-least-fair SAH into one category which takes value one. On the full sample, 93% of answers are reporting at least as fair health. The prevalence of at-least-fair SAH ranged from 52% (Croatia) to Ireland (98%). Baltic states respondents tend to report low values and more generally, countries of central or eastern Europe. The tendency to report at-least-fair SAH may differ between countries due to cultural factors (see for instance, Babones (2009) or Reile et al. (2014)). Alternatively, it may reflect poor health problems are widespread. Since we cannot disentangle the two effects, we avoid commenting on the difference in average SAH.

EU-SILC also contains information on sources of biological health variation (notably age and gender) and social sources of inequality. Age represents a source of unavoidable degradation in health conditions in every country. We assume that there is some common pattern in Europe about the deterioration of health with aging. In a preliminary step, we introduce fixed effects of belonging to one of four age classes: 25-30, 30-39, 40-49, 50-59. By cleaning in that way the age effect, we also clean the correlation of circumstances with age at the European level but not at the country level. So, it would not affect the comparison of *IOpH* across European countries.

In the set of circumstances suffered by individuals, we have included sex and variables describing the socioeconomic background of all respondents. These variables relate to the immigration status (whether native, first or second generation immigrant) and the family situation of the respondent when she was around 14 years old. These questions concern immigration history, education, the economic activity and the housing ownership of the respondent's parents. We consider the majority of items as potentially relevant circumstances beyond individual control.[4]
Two pairs of questions concern parents' country of birth and citizenship (born in the respondent's present country of residence, born in another EU-27 country, born in another European country, born outside Europe).
Father and mother's educations are categorized according to the International Standard Classification of Education 1997 (ISCED-97). Education attainments are aggregated into four groups: Illiterate, Low, Medium, High. Low corresponds to levels 0, 1, and 2 of ISCED-97 (except the persons who are illiterate). Medium corresponds to levels 3 and 4, and High level of education corresponds to levels 5 and 6 of ISCED-97. To these groups, we add a group that includes all respondents with an unknown father/mother.

[3] The latter category is not considered in the analysis.
[4] We did not include all possible observable circumstances because the number of parameters necessary to estimate a latent class model are proportional to the number of circumstances and then number of levels they can take. It is therefore necessary to ponder the inclusion of controls to obtain a sufficient number of degrees of freedom.

Father and mother's main occupation qualifications are described by an ordinal variable coded according to the ISCO-08 (COM) classification (International Standard Classification of Occupations, published by the International Labour Office). Managers, Professionals, Technicians, Clerical support workers, Service and sales workers, Skilled agricultural, forestry and fishery workers, Craft and related trade workers, Plant and machine operators, and assemblers, Elementary occupations. To these categories, we add a group that includes all respondents whose father/mother did not work, was unknown or was dead.
Descriptive statistics for the main categories of all circumstances are reported in the Appendix (Tables 1A and 2A). Table 3A reports the percentage of missing information for each circumstance. The prevalence of missing information is very heterogeneous. We warn the reader to consider with caution estimates of countries with a high prevalence of missing information.[5]

3.2 Inside the black-box of latent type: an example

For a low number of types it is possible to learn something about how latent groups are obtained looking at the covariance of observable circumstances and latent types' membership. To illustrate, we look at Portugal which belongs to the top countries in terms of *IOpH*. Without getting into the details of all the circumstances considered, Table 2 and Table 3 display the share of individuals belonging to each of the three latent types disaggregated by father occupation and mother education.

Table 2: Latent type membership by father occupation (ISCO-08) (Portugal -- 3 latent types)

Father occupation	Type 1	Type 2	Type 3
Father dead/unknown/not working	29.30%	17.80%	52.90%
Elementary	27.40%	4.20%	68.40%
Plant Operator	76.20%	6.40%	17.40%
Craft/Trades	61.20%	4.80%	34.00%
Agriculture	18.90%	3.90%	77.20%
Service	85.00%	5.20%	9.80%
Clerical	77.50%	19.30%	3.20%
Technician	80.10%	16.80%	3.10%
Professional	22.50%	77.50%	0.00%
Manager	69.00%	26.90%	4.10%

Source: EU-SILC, 2011

Table 3: Latent type membership by mother education (Portugal -- 3 latent types)

Mother education	Type 1	Type 2	Type 3
Illiterate	11.50%	4.10%	84.40%
Low	74.90%	6.70%	18.40%
Medium	25.80%	71.10%	3.10%
High	0.00%	100.00%	0.00%

Source: EU-SILC, 2011

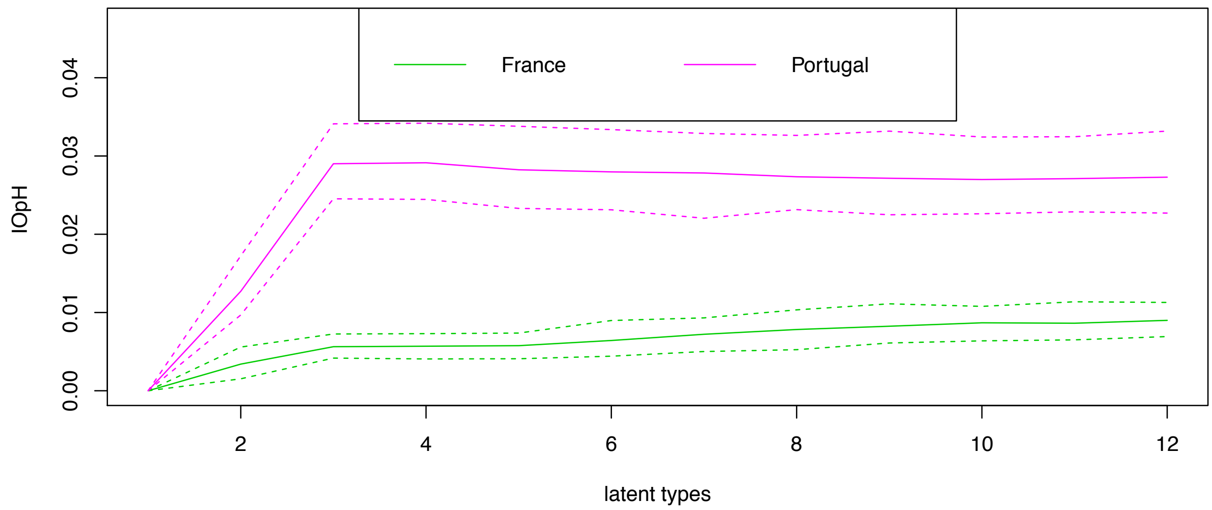
[5] For example, all `register countries', (Denmark, Finland, Iceland, the Netherlands, Norway, Sweden and Slovenia) in which only a selected household respondent receives a personal questionnaire and household and income variables are collected either through a register or through the selected respondent.

Type 3 is made of individuals with a low level of occupation and low level of mother education. Not surprisingly, individuals belonging to type 3 have the lowest average level of self-assessed health. On the opposite, individuals with a father employed in intermediate or high occupations are registered in Type 1. 69% of the respondent declaring to have a father working as a manager belong to type 1. The level of education of the mother for this type tends instead to be low. This type has a level of predicted health close to the population average. Finally, type 2 is to a large extent, made of individuals with a father with a medium-high level of occupation but with a highly educated mother. All respondents reporting a mother with a high level of education belong to type 2. This type has the highest level of age-adjusted SAH in the sample. Father education and mother occupation tend to describe a similar picture. Looking inside the black-box of latent types, we obtain a picture of how parental education and parental occupation interact in shaping health opportunities in Portugal.

3.3 IOpH Dominance

In this section, we present the results about dominance of opportunity curves. The opportunity curve plots the dissimilarity index calculated on the counterfactual distribution against the number of latent types. Figure 2 shows the health opportunity curves for France and Portugal. In this case, there is a clear dominance of the former over the latter. No matter what number of latent types we consider, France is estimated to have a much lower level of inequality of opportunity than Portugal. The difference is statistically significant in all cases.

Figure 2 Opportunity-inequality curve: Portugal and France



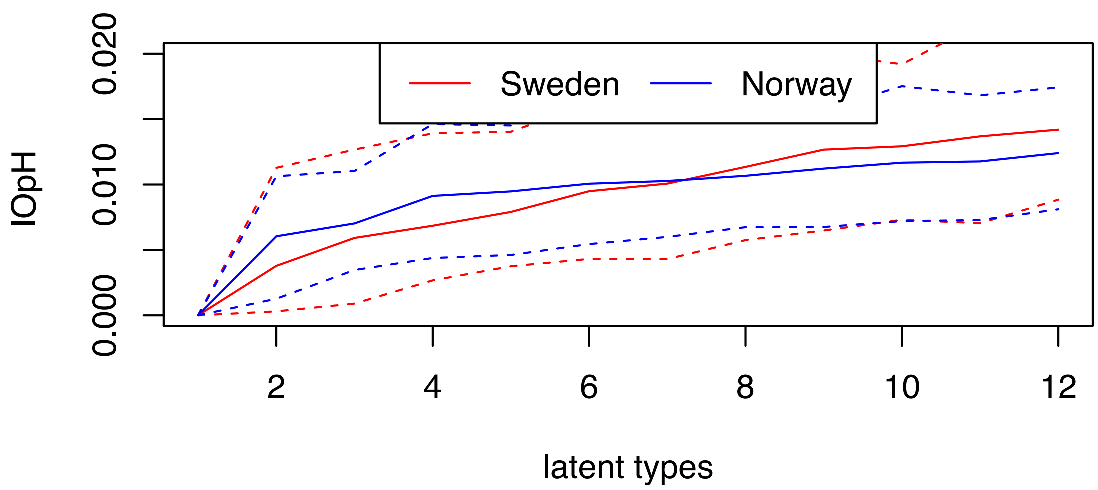
Source: EU-SILC, 2011. Confidence intervals are based on the 2.5-th and 97.5-th centile of the distributions of 200 bootstrap replications of $IOpH_l$ for each number of latent types.

Portugal shows a particularly steep opportunity-inequality curve for a low number of latent types. The level of inequality that can be explained by three latent types is already around the inequality explained by 12 types. In other countries, the inequality explained by between-type inequality grows more gradually, France making no exception. The shape of the Portuguese opportunity curve recalls back our attention to the partition described in section 3.2. The tree types considered do not only have

a particularly intuitive structure, but they are also hugely significant in explaining inequality in health in Portugal.
On a more technical note, although the two curves tend to be increasing, we can notice examples of local decreasingness (for instance, the curve in Portugal is flat or decreasing for a number of latent types larger than three). On average, a larger number of types translates into a higher between-type variance. However, nothing guarantees that the two curves are monotonically increasing. This feature is an important point. Individuals are assigned to latent types to minimize the dependence of observable circumstances within types. It ensures that a larger number of types will reduce the within-type correlation of observable circumstances. However, this does not automatically translate into an increase in the between-group variance. Ferreira and Gignoux (2011) prove that a partition based on a larger number of types (whatever adopting a parametric or a non-parametric approach) necessarily increases the estimated inequality of opportunity. This reasoning cannot be extended for the latent-type approach because the set of latent types for *n* latent types is not a thinner partition of the set of latent types for *n-1* latent types.

As shown in Figure 3 for Sweden and Norway, health inequality of opportunity curves can intersect. In the specific case *IOpH* appears higher in Norway when the number of types is smaller than seven, then *IOpH* is higher in Norway for the partitions in types with larger number of types. When inequality of opportunity curves do intersect, or when their vertical distance is never statistically significant, it is impossible to rank the two countries independently from the number of classes assumed. The two viable alternatives consist of choosing a particular number of latent types or aggregating information contained in the opportunity curves in a summary index. Note, however, that for large numbers of latent types, the confidence interval for both curves tends to become large. Indeed, the approach needs more parameters to be estimated and is degree-of-freedom consuming. A fine-grained partition in latent types is obtained at the cost of increased uncertainty about their sign and magnitude. A trade-off between robustness and sharpness is unavoidable.

Figure 3: Inequality of opportunity dominance: Sweden and Norway

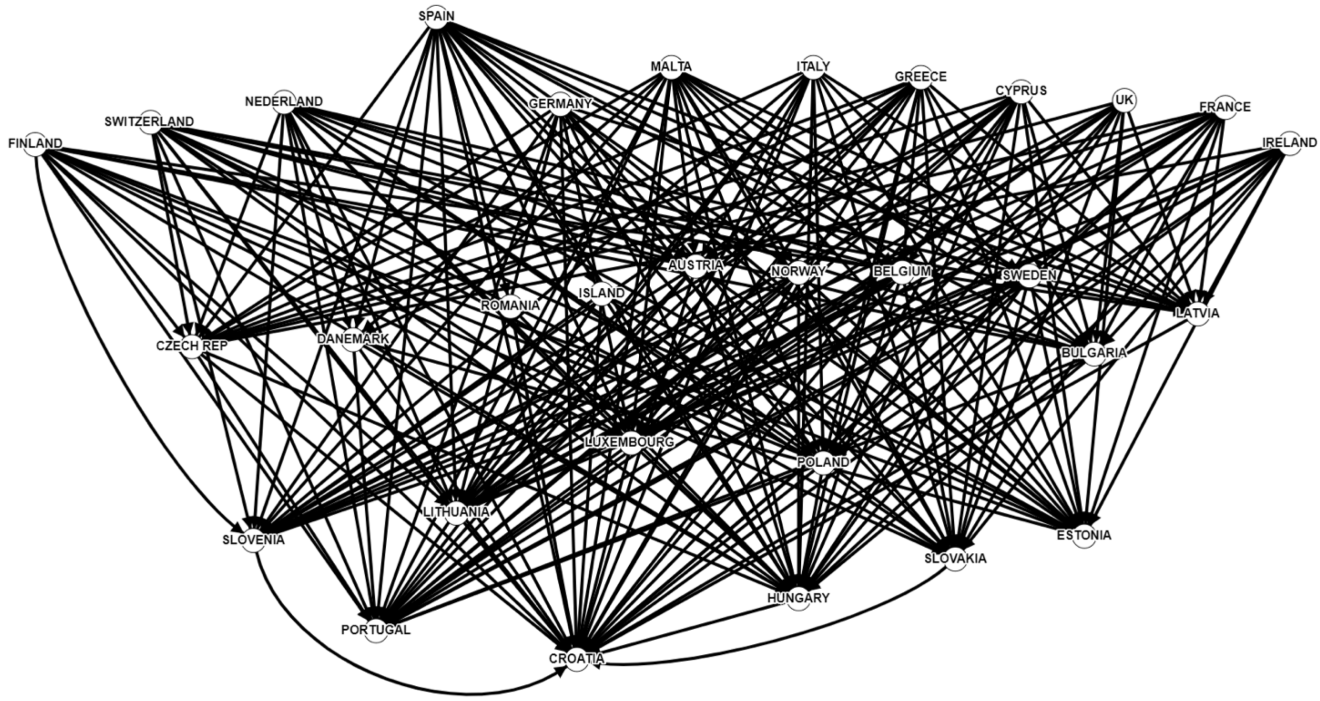


Source EU-SILC, 2011. Confidence intervals are based on the 2.5-th and 97.5-th centile of the distributions of 200 bootstrap replications of $IOpH_l$ *for each number of latent types.*

Table 4A in the Appendix reports dominance in terms of *IOpH* when up to 12 latent types are considered. The cell in row *i* and column *j* reporting > should be read as: country *i* has larger *IOpH* than country *j*. An empty cell means that the two opportunity curves intersect at least once.

Of the 465 possible pairwise comparisons in our 31-country sample, we find dominance in 375 pairs. As signaled by stars, out of the 465 comparisons, 227 are statistically significant at least at 90%, which represents about 50% of all possible comparisons. The price to pay for getting robustness is relatively moderate in terms of incompleteness.
It is possible to rank many countries unambiguously against most other countries as shown by the directed graph in Figure 4 corresponding to Table 4A.

Figure 4. The directed graph of dominance according to inequality of opportunity (The top countries are those with lower *IOpH*)



Source EU-SILC, 2011.

Formally, we find 5 equivalence classes. An equivalence class is defined as no relation of dominance (at least at 10%) between the elements of the class: each element is not dominated by other elements of the subset considered (at 10%). The ranking of each equivalence class is obtained sequentially by removing the set of non-dominated countries at the further step. To obtain a thiner grain ranking, we choose to rank countries of the same equivalence class according to the number of dominated countries.

The top-cycle set – the countries that are undominated in the whole set – are Western European countries, with all Mediterannean countries except Portugal. In this top-cycle set comprising 12 countries, Spain comes first when counting the number of dominated countries, 19. Indeed, Spain dominates all countries not belonging to the top-cycle set. Then Malta comes next with 17 dominated countries. Greece and the Netherlands closely follow with 15 dominated countries and then Cyprus and Switzerland with 14 dominated countries. When removing these 12 countries, we find another group of 10 countries which are undominated among the remaining countries. They are mainly Western European countries with some Estearn European countries such as the Czech Republic, Romania, Bulgaria and Latvia. Quite surprinsingly, all Nordic countries belong to this group. Next come a group of 6 Eastern European countries plus Luxembourg. Finally, a pack of three countries compose the bottom-cycle set, Hungary a little bit ahead, and then Portugal and Croatia which appears as the worst country according to equality of health opportunity with 28 dominating countries out of 31.

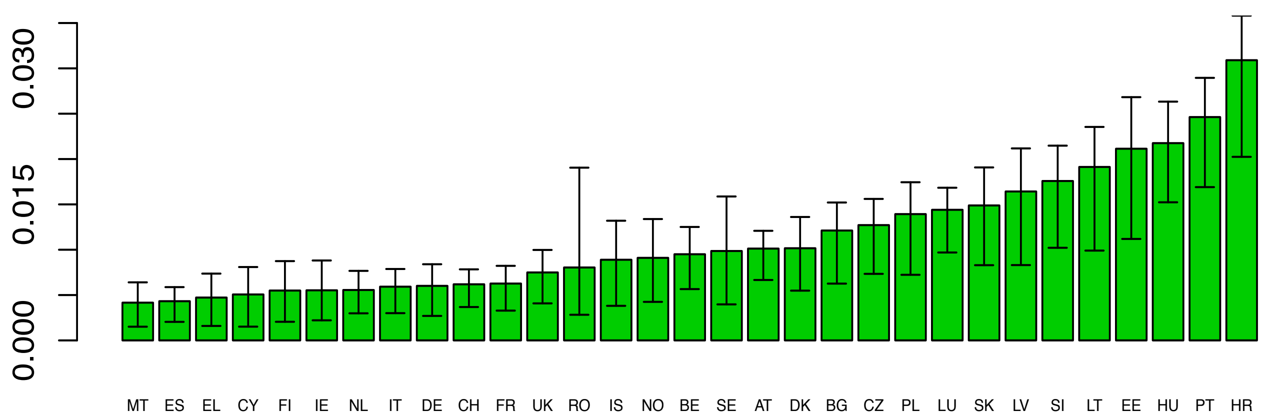
So the robust ranking of countries is the following:
Rank 1: Spain, Malta, Greece, Nederland, Cyprus, Switzerland, Finland, France, Italy, Ireland, United Kingdom, Germany,
Rank 2: Belgium, Norway, Sweden, Island, Austria, Denmark, Czech Republic, Bulgaria, Romania, Latvia
Rank 3: Luxembourg, Poland, Lithuania, Slovenia, Slovakia, Estonia
Rank 4: Hungary
Rank 5: Portugal, Croatia

To sump up, there is a consistent pattern with Western European countries being almost always better ranked than Eastern European countries, except Portugal the worst performer with Croatia.

3.4 IOpH indexes

In order to obtain a complete ranking of countries, we calculate $IOpH$ as the mean of the 12 possible values — this summary index averages across all possible latent model specifications between-type dissimilarity. Figure 5 presents the ranking of countries according to $IOpH$.The picture is consistent with what is returned by Figure 4. Worst performing countries are concentrated among Eastern European countries with the notable exception of Portugal. Countries with the lowest *IOpH* are instead heterogeneous. Including countries of the Mediterranean area, the two English-speaking countries, Nordic and continental countries.

Figure 5: Inequality of health opportunity in Europe



Source: EU-SILC, 2011. Confidence intervals are based on the 2.5-th and 97.5-th centile of the distributions of 200 bootstrap replications of IOpH.

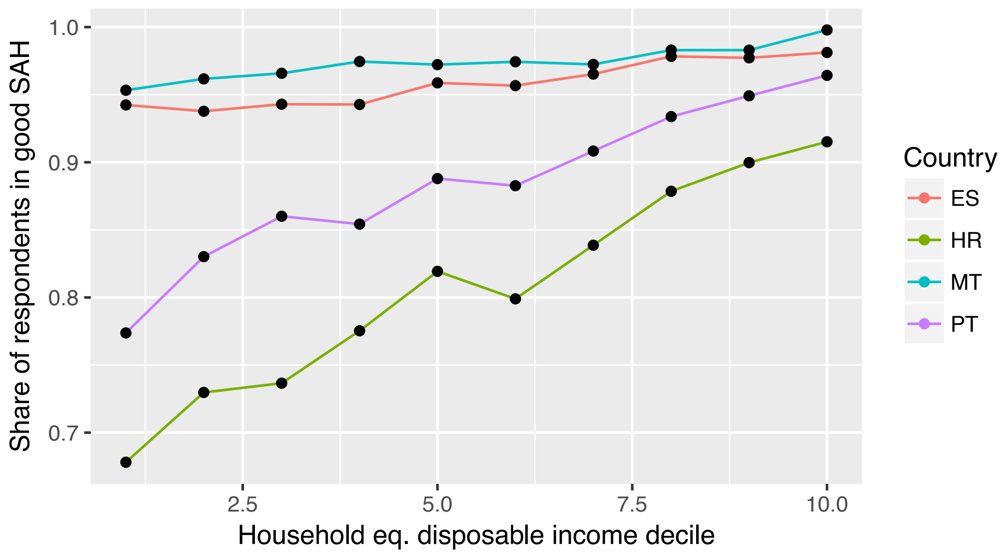
5. Discussion

There is a couple of issues that should be discussed. First, one can ask whether our results depend much on our original way to partition the population into types. To assess to what extent our ranking is the result of our measurement approach, we replicate the standard methodology proposed by Ferreira and Gignoux (2011) and Figure 8A in the Appendix shows the high – but not perfect - correlation between the ranking obtained with the two methods. We interpret positively this result showing that the obtained ranking is really what the data convey.

Second, the SAH may be subject to misreporting and is associated with comparability problems at both the individual and between countries (eg., Crossley and Kennedy 2002 and Bago d'Uva et al. 2008 and 2009). The reliability of the SAH values in EU-SILC may be questioned for a few countries. We deal with this issue from three various perspectives, the link with income, the stability of responses to SAH questionnaire, and the external validity.

The first perspective is the almost independence relation of SAH with income in some countries. When we plot the average binary SAH variable per income decile of the respondents for two top countries (Malta and Spain) and the two bottom countries (Portugal and Croatia), it appears that the socioeconomic gradient for the two top countries is rather flat. (See the Appendix Figure 7A)

Figure 6: self-assessed health status by income decile



Source: EU-SILC, 2011.

Au and Johnston (2014) find that some health dimensions - especially vitality - are consistently crucial to an individual when they assess their health, while other dimensions are inconsequential. They demonstrate that this fact provides insight into why some studies find weak income gradients in SAH. Instrumental-variable regression results show that shocks to household income do not affect SAH.

Indeed, this weak association of self-assessed health with socioeconomic backgrounds in Greece is known (Daniilidou et al., 2004, Eurostat, 2017), on the other hand, it is likely to have been rising in the years following 2011. See Effie and Koutsogeorgou (2014) for a review of the consequences of the Great Recession on health in Greece.

Another perspective is the stability of the results for the same country over time with different surveys. Crossley and Kennedy (2002) found that more than 1/ 4 of respondents change their reported health status before and after an additional set of health-related questions. It can, therefore, be illuminating to compare the SAH values for EU-SILC with values obtained in older surveys. Table 4 compares the proportion of respondents who reports less-than-good health obtained with EU-SILC with two other sources. National surveys compiled by Hu and al. (2017) from mid-1990 to 2010 at the latest for an age range older than ours, 30-79, and waves 3 and 4 from World Value Survey when interviews were conducted between 1994 and 2004 compiled by Babones (2009). In all cases, it is age-adjusted health. Hu and al. ((2017) report results where age is standardized to the European Standard population using the direct standardization method. Babones (2009) has benchmarked the data to the SAH of a 40-year old male.

Table 4: Comparison of less-than-good SAH average per country according to three sources.

	EU-SILC	National Surveys Men	National Surveys Woman	World Value Survey
Austria	26	28-31	29-33	
Belgium	21	23-25	27-31	
Bulgaria	19			43
Switzerland	15			17
Cyprus	18			
Czech Re	29	41-70	47-73	47
Germany	28			41
Denmark	26	19-22	23-27	
Estonia	48	58-65	57-67	61
Greece	13			
Spain	16	27-31	38-41	28
Finland	13	36-44	33-42	28
France	27			
Croatia	63			49
Hungary	38			53
Ireland	13	15-17	12-16	
Iceland	16			
Italy	25	37-45	45-53	
Lithuania	53	54-66	51-68	55
Luxembourg	25			
Latvia	49			64
Malta	21			
Netherlands	19	26	30	
Norway	22			21
Poland	36	40-64	44-71	59
Portugal	44	50-62	64-77	
Romania	19			47
Sweden	16	21-23	26-28	23
Slovenia	34			54
Slovakia	32			
Britain	20	34-40	40-46	
Scotland		24-27	25-26	

Source: Column 1: EU-SILC 2011. Column 2, Table 1 Hu et al. (2017), Column 3: Table 1 Babones (2009).

We will not emphasize that there are many reasons (age group, period) why the results may be volatile. What is quite noticeable is that for a couple of Western European Countries, they are not. Ireland is one of them, and it is quite remarkable since it is one of our best performers according to our ranking. Austria, Belgium, Sweden also show stable results. Poland in Eastern Europe is also fairly stable but for levels, unfortunately, lower of SAH. Romania is not and exhibits an improving situation quite spectacular as well as Britain.

A third perspective is the external validity of the SAH results. There are other questions in EU-SILC about health and specifically self-reported chronic illness. It can be viewed as quite odd to declare a chronic illness and to report good or very good health. However, it is not so rare in the full sample as shown by the following distribution. About 30 percent of the respondents that declare to have chronic illness also declare to be in good or very good health condition (Table 5).

Table 5: SAH and chronic illness

	Yes	No	Do not know
Very good	3.85%	27.10%	5.18%
Good	27.37%	59.04%	44.37%
Fair	42.62%	13.07%	43.69%
Bad	21.31%	0.70%	6.76%
Very bad	4.85%	0.08%	0.00%

Source: EU-SILC 2011

Interestingly, *IOpH* shows a weak correlation with other significant countries' health indicators and levels of inequality or inequality of opportunity, as shown by Table 6. The indicators are reported in Table 5A in the Appendix. The majority of correlations are rather weak, only in two cases they are statistically significant.

The age-adjusted health, calculated from EU-SILC, is the share of respondents whose self-assessed health is very good or good. This measure of average outcome is the indicator most clearly correlated with *IOpH* ($r=-0.89$, $p\text{-value}<0.01$). The first scatterplot (Figure 6A) shows to the top-left a cluster of countries with a high level of health inequality and low level of self-reported health. They are former transition economies and Portugal. This negative correlation can be understood as average and equality goes hand in hand in health matters. Interestingly, this statement seems also true for average income and *IOpH* ($r=-0.45$, $p\text{-value}<0.05$).

Table 6: Correlation of IOpH with other health and inequality indicators

	IOpH	Age adj. SAH	Health expenditure	Inequality (income)	IOp (income)	Avg. Eq. Disp. Income
IOpH	1					
Age adj. SAH	-0.8908***	1				
Health expenditure	-0.2067	0.2586	1			
Inequality (income)	0.2004	-0.2147	0.2041	1		
IOp (income)	0.2118	-0.0997	0.2981	0.7587***	1	
Avg. Eq. Disp. income	-0.4498**	-0.2088	-0.0306	-0.2736	0.2586	1

*Source: see Table 7 in the Appendix. Statistical significance: ***=0.01, **=0.05, *=0.1.*

6. Conclusions

We have proposed a robust method for ranking populations in terms of inequality of health opportunity. Using the 2011 round of the EU-SILC survey, we have been able to estimate inequality of opportunity in health in 31 countries. In order to implement such a measure, we identify health outcome as reporting “good” or “very good” general health condition. This choice makes our analysis comparable to other contributions that have generally focused on self-assessed health conditions. Circumstances beyond individual control that are considered sources of unequal opportunities are also in line with the existing literature.

To rank countries in terms of inequality of health opportunity, we extend the approach recently introduced by Li Donni et al. (2015). After removing variability due to age, we estimate a latent class model to identify latent types, that is groups of individuals exposed to a similar mix of observable and unobservable circumstances beyond individual control. Contrary to what has been suggested by Li Donni et al. (2015), we do not select a predetermined number of latent types to perform cross-country comparisons; we instead sequentially measure inequality of opportunity for a different number of latent groups. In doing so, we introduce a dominance procedure to rank populations in terms of inequality of opportunity. Our proposal completes what has been suggested by Li Donni et al. (2015) in making comparisons in terms of inequality of opportunity insensitive to the choice of an exact number of unknown types. Dominance in terms of equality of opportunity can be shown by plotting the level of inequality of opportunity measured for each possible number of latent types. We have called this graph the *opportunity-inequality curve* because it shows how inequality explained by latent types' membership changes when the population is partitioned in an increasing number of latent types.

The *opportunity-inequality curve* is a partial criterion to rank countries. When two curves intersect, it is impossible to rank them in terms of equality of opportunity. If one is interested in a complete ordering of countries, there is a direct way to aggregate information contained in the curve by taking the average of its coordinates. This summary index allows us to rank any pairs of countries as more, less or unequal in terms of inequality of opportunity in health.

Implementing this approach, we exhibit statistically significant dominance in almost 1/2 of pairwise comparisons. Mediterranean countries, except Portugal, perform well, contrary to Eastern European countries. Moreover, using the summary index of inequality of opportunity in health, we can rank all countries in the sample and correlate our estimates with other important indicators of health and inequality. Inequality of health opportunity appears to be strongly and negatively correlated with age-adjusted self-reported health condition. Thus, in terms of association, health equality and good health go hand in hand at least in Europe.

References

- Arneson, R., 1988. Equality of Opportunity for Welfare, Philosophical Studies, 56: 77-93.
- Au, N. and Johnston DW (2014) Self-assessed health: what does it mean and what does it hide? Social Science & Medecine 121, 21-28.
- Babones SJ. 2009 The consistency of self-rated health in comparative perspective. Public Health 123:199–201
- Bago d'Uva, T. and Jones, A.M. 2009 "Health care utilisation in Europe: new evidence from the ECHP", Journal of Health Economics 28, 265-279.
- Bago d'Uva, T., Jones, A.M. and van Doorslaer, E. 2009 "Measurement of horizontal inequity in health care utilisation using European panel data", Journal of Health Economics, 28, 280-289.
- Björklund, A., Markus, J., Roemer, J.E., 2011. Equality of Opportunity and the Distribution of Long-Run Income in Sweden. Social Choice and Welfare, 39 (2-3): 675-696.
- Bourguignon, F., Ferreira, F., Menendez, M., 2007. Inequality of Opportunity in Brazil, Review of Income and Wealth, 53(4): 585-618.
- Bricard, D., Jusot, F., Trannoy, A., Tubeuf, S., 2013. Inequality of Opportunity in Health and the Principle of Natural Reward: Evidence from European Countries, in: P. Rosa Dias and O. O'Donnell (eds.) Health and Inequality, Research on Economic Inequality, Volume 21, Emerald Group Publishing Limited: 335-370.
- Brunori, P., Peragine V., Serlenga L., 2019, Upward and downward bias when measuring inequality of opportunity. Social Choice and Welfare, Volume 52, Issue 4, pp 635-661.
- Brunori, P., Hufe, P., Mahler, D. G., 2018. "The roots of inequality: Estimating inequality of opportunity from regression trees", Policy Research working paper; no. WPS 8349. Washington, D.C. : World Bank Group.
- Burström, K., Johannesson, M., Diderichsen, F., 2005. Increasing socio-economic inequalities in life expectancy and QALYs in Sweden 1980-1997, Health Economics, 14: 831-850.
- Carrieri, V., Jones, M.A., 2016. Inequality of opportunity in health: A decomposition-based approach", HEDG Working Papers 16/5, University of York.
- Carrieri, V, Davillas A, A. M. Jones, 2019, A latent class approach to inequity in health using biomarker data, ISER WP 2019-09.
- Checchi, D., Peragine, V., 2010. Inequality of opportunity in Italy, Journal of Economic Inequality, 8: 429-450.
- Cohen, G.A., 1989. On the Currency of Egalitarian Justice, Ethics, 99: 906-944.
- Crossley, T, S. Kennedy 2002. The reliability of self-assessed health status. Journal of Health Economics. Volume 21, Issue 4, July 2002, Pages 643-658
- Daniilidou N.V., Gregory S., Kyriopoulos J., Zavras D.J., 2004. Factors associated with self-rated health in Greece: a population-based postal survey. European Journal of Public Economics. 14 (2): 209-211.
- Dworkin, R., 1981. What is equality? Part 2: Equality of resources, Philosophy and Public Affairs, 10: 283-345.
- Effie S. and Koutsogeorgou E., 2014. Effects of the economic crisis on health and healthcare in Greece in the literature from 2009 to 2013: A systematic review. Health Policy. 115 (2–3): 111-119.

- Eurostat, 2017. Self-perceived health by sex, age and income quantile. Available at: http://ec.europa.eu/eurostat access: 28/10/2017.
- Ferreira, F.H.G., Gignoux, J., 2011. The measurement of inequality of opportunity: theory and an application to Latin America, The Review of Income and Wealth, 57(4): 622-657
- Fleurbaey, M., 2008. Fairness, responsibility and welfare, Oxford University Press.
- Garcia-Gomez, P., Schokkaert, E., Van Ourti, T., Bago d'Uva, T., 2012. Inequity in the Face of Death", CORE Working Paper no. 2012/14.
- James G., Witten D., Hastie T., Tibshirani R. (2014). "An Introduction to Statistical Learning with Applications in R", Springer, New York.
- Jusot, F., Tubeuf, S., Trannoy, A., 2013. Circumstances and efforts: how important is their correlation for the measurement of inequality of opportunity in health?", Health Economics, 22: 1470-1495.
- Kunst AE, Bos V, Lahelma E, et al. Trends in socioeconomic inequalities in self-assessed health in 10 European countries. Int J Epidemiol 2005;34: 295–305
- Lantz, P.M., Lynch, J.W., House, J.S., Lepkowski, J.M., Mero, R.P., Musick, M.A., Williams, D.R., 2001. Socioeconomic disparities in health change in a longitudinal study of US adults: the role of health-risk behaviors, Social Science & Medicine, 53(1): 29-40.

- Lanza S., Xianming T., Bethany B., 2013. Latent Class Analysis With Distal Outcomes: A Flexible Model-Based Approach. Structural Equation Modeling, 20(1): 1-26.
- Lazar, A., 2013. Ex-ante and Ex-post Measurement of Inequality of Opportunity in Health: Ev- idence from Israel, in: P. Rosa Dias and O. O'Donnell (eds.) Health and Inequality, Research on Economic Inequality, Volume 21, Emerald Group Publishing Limited: 371 - 395.
- Li Donni, P., Peragine, V., Pignataro, G., 2014. Ex-ante and ex-post measurement of equality of opportunity in health: a normative decomposition", Health Economics, 23: 182 -198.
- Li Donni, P., Rodriguez, J.G., Rosa Dias, P., 2015. Empirical definition of social types in the analysis of inequality of opportunity: a latent classes approach, Social Choice and Welfare, 44(3): 673-701.
- Mackenbach, J.P., Johan, P., Stirbu, I., Albert-Jan, R.R., Maartje, M.S., Menvielle, G., Mall, L., Kunst, A.E., 2008 Socioeconomic Inequalities in Health in 22 European Countries, New England Journal of Medicine, 358 (23): 2468-2481.
- Marmot Review, 2010. Fair Society, Healthy Lives: Strategic Review of Health Inequalities in England Post 2010, Marmot Review, London.
- Paes de Barros R., F H.G., Ferreira, J.R., Molinas Vega, JS., Chanduvi, 2009. Measuring Inequality of Opportunities in Latin America and the Caribbean, World Bank Publications, n° 2580.
- Ramos, X. and D. Van de Gaer 2016. Approaches to inequality of opportunity: principles, measures and evidence. Journal of Economic Surveys 30(5), 855–83.
- Roemer, J.E., 1998, Equality of Opportunity, Harvard University Press, Cambridge, Ma.
- Roemer, J.E., Trannoy, A., 2015. Equality of Opportunity, Handbook of Income Distribution.
- Rosa Dias, P., 2009. Inequality of opportunity in health: evidence from the UK cohort study", Health Economics, 18: 1057-1074.
- Rosa Dias, P., 2010. Modelling opportunity in health under partial observability of circumstances, Health Economics, 19: 252-264.
- Reile R, Helakorpi S, Klumbiene J, et al. The recent economic recession and

self-rated health in Estonia, Lithuania and Finland: a comparative cross-sectional study in 2004–2010. J Epidemiol Community Health 2014;68:1072–9.

- Toch-Marquardt, M., Menvielle, G., Eikemo, T.A., Kulh`anov`a, I., Kulik, M.C., Bopp, M., Es- naola, S., Jasilionis, D., M¨aki, N., Martikainen, P., Regidor, E., Lundberg, O., Mackenbach, J.P. Euro-GBD-SE consortium, 2014. Occupational Class Inequalities in All-Cause and Cause-Specific Mortality among Middle-Aged Men in 14 European Populations during the Early 2000s, PLoS One, 9(9): e108072.
- Trannoy, A., Tubeuf, S., Jusot, F., Devaux, M., 2010. Inequality of opportunities in health in France: a first pass", Health Economics, 19(8): 921-938.
- Van De Gaer, D., Vandenbossche, J., Figueroa, J.L., 2012. Children's Health Opportunities and Project Evaluation: Mexico's Oportunidades Program", CORE DP 2012/15.

Appendix – Additional tables

Table 1A: Descriptive statistics of circumstances (1)

								Father birth		Mother birth		Father edu.		Mother edu.	
Country	sample	SAH	Adj. SAH	Change	Female	Native	Immingrant (EU)	Native	Immigrant (EU)	Native	Immigrant (EU)	Low	Medium	Low	Medium
AT	5973	0.931	0.932	-0.08%	0.499	0.7895	0.069	0.7529	0.0927	0.7473	0.0975	0.4048	0.4318	0.5897	0.3325
BE	4178	0.927	0.926	0.12%	0.503	0.8417	0.071	0.8089	0.1002	0.8105	0.1018	0.5171	0.2168	0.5768	0.2189
BG	5507	0.95	0.95	-0.03%	0.504	0.9958	0.001	0.9918	0.0041	0.9915	0.0033	0.5031	0.3722	0.4859	0.3731
CH	6410	0.966	0.966	0.00%	0.497	0.6856	0.197	0.5895	0.2881	0.5686	0.3075	0.2296	0.4953	0.4098	0.4039
CY	4459	0.961	0.959	0.13%	0.477	0.7974	0.092	0.8170	0.0789	0.8165	0.0772	0.6741	0.1801	0.6883	0.1639
CZ	6017	0.92	0.92	-0.06%	0.435	0.9636	0.027	0.9175	0.0677	0.9203	0.0640	0.6472	0.2281	0.6761	0.2697
DE	10473	0.936	0.939	-0.36%	0.506	0.8811	0	0.8132	0.1868	0.8221	0.1779	0.1240	0.5369	0.2896	0.5089
DK	2107	0.918	0.918	-0.04%	0.505	0.9267	0.025	0.9243	0.0263	0.9172	0.0310	0.3535	0.4282	0.5397	0.2831
EE	4272	0.775	0.776	-0.18%	0.465	0.8604	0	0.7053	0.2947	0.7052	0.2948	0.3517	0.3965	0.3364	0.4058
EL	5704	0.966	0.965	0.09%	0.5	0.8956	0.024	0.8973	0.0153	0.8985	0.0154	0.5975	0.1370	0.5985	0.1326
ES	14514	0.964	0.962	0.21%	0.505	0.8436	0.046	0.8463	0.0427	0.8452	0.0419	0.7784	0.0646	0.8134	0.0483
FI	2467	0.942	0.943	-0.09%	0.528	0.9405	0.021	0.9467	0.0126	0.9512	0.0098	0.5435	0.2261	0.5300	0.2539
FR	9959	0.939	0.939	-0.04%	0.487	0.8863	0.036	0.8013	0.0778	0.8166	0.0678	0.7248	0.0768	0.7433	0.0803
HR	5879	0.525	0.525	-0.02%	0.49	0.8789	0.016	0.8304	0.0061	0.8325	0.0072	0.5127	0.3358	0.6855	0.1951
HU	11951	0.886	0.895	-1.06%	0.481	0.9888	0.008	0.9774	0.0181	0.9793	0.0169	0.6243	0.2564	0.6664	0.2497
IE	3067	0.978	0.978	-0.02%	0.425	0.7969	0.136	0.8071	0.0997	0.8029	0.1044	0.5842	0.2621	0.5443	0.3326
IS	1446	0.949	0.949	-0.01%	0.511	0.9106	0.055	0.9245	0.0482	0.9136	0.0565	0.3284	0.4960	0.6250	0.2783
IT	18662	0.92	0.919	0.10%	0.5	0.8883	0.038	0.9060	0.0245	0.9015	0.0260	0.7785	0.1478	0.8192	0.1151
LT	4356	0.836	0.838	-0.28%	0.471	0.9426	0.004	0.9322	0.0035	0.9378	0.0018	0.6081	0.2530	0.5390	0.3210
LU	6391	0.931	0.929	0.15%	0.499	0.4866	0.4	0.3983	0.4784	0.3844	0.4958	0.4931	0.3291	0.5916	0.2522
LV	4933	0.9	0.901	-0.17%	0.476	0.8696	0	0.6990	0.3010	0.7168	0.2832	0.4762	0.3675	0.4230	0.4183
MT	4017	0.978	0.978	-0.04%	0.505	0.9488	0	0.9688	0.0312	0.9640	0.0360	0.5735	0.1822	0.6628	0.1444
NL	5293	0.95	0.951	-0.07%	0.487	0.8900	0.018	0.8684	0.0272	0.8666	0.0268	0.3922	0.3002	0.5452	0.2933
NO	2242	0.931	0.931	-0.09%	0.524	0.9159	0.04	0.9144	0.0442	0.8977	0.0474	0.3274	0.3886	0.3658	0.4425
PL	12476	0.911	0.914	-0.32%	0.49	0.9987	4E-04	0.9674	0.0128	0.9703	0.0112	0.4739	0.4523	0.5317	0.4074
PT	5634	0.898	0.896	0.24%	0.491	0.9087	0.022	0.9435	0.0065	0.9401	0.0082	0.7138	0.0313	0.6370	0.0290
RO	5273	0.947	0.943	0.50%	0.508	0.9987	0	0.9969	0.0013	0.9976	0.0011	0.8348	0.0997	0.8221	0.1240
SE	466	0.966	0.966	0.00%	0.523	0.8839	0.046	0.8330	0.0739	0.8245	0.0821	0.5269	0.1574	0.5083	0.2219
SI	4560	0.906	0.906	-0.11%	0.496	0.8667	0	0.7959	0.2041	0.8169	0.1831	0.7092	0.1735	0.7725	0.1530
SK	5864	0.909	0.912	-0.31%	0.477	0.988	0.01	0.976	0.021	0.973	0.024	0.385	0.526	0.464	0.487
UK	5712	0.951	0.95	0.09%	0.472	0.855	0.04	0.812	0.063	0.819	0.063	0.516	0.241	0.685	0.106

Source: EU-SILC, 2011. Averages are calculated using sampling weights. Health is the share of respondents reporting good or very good general health condition; Age-adjusted Health is obtained by direct standardization based on the population structure of the entire sample.

Table 2A: Descriptive statistics of circumstances (2)

	Father occupation				Mother occupation				
Country	Elementary	Plant Operat	Craft/Trades	Agriculture	Not working	Elementary	Agriculture	Service	Tenancy
AT	0.0850	0.0637	0.2862	0.2862	0.4637	0.0857	0.1304	0.1546	0.5900
BE	0.0441	0.1486	0.2276	0.2276	0.6459	0.0698	0.0017	0.0449	0.7725
BG	0.1463	0.2204	0.2325	0.2325	0.0793	0.1479	0.1939	0.1441	0.9235
CH	0.0525	0.0776	0.2248	0.2248	0.4719	0.0666	0.0531	0.1243	0.5496
CY	0.1259	0.1233	0.2465	0.2465	0.5102	0.2192	0.0367	0.0681	0.7855
CZ	0.0564	0.2086	0.3236	0.3236	0.0834	0.1378	0.0764	0.1608	0.6168
DE	0.0392	0.1534	0.2625	0.2625	0.4845	0.0337	0.0270	0.1171	0.5204
DK	0.0100	0.0690	0.2908	0.2908	0.3392	0.0018	0.0368	0.2169	0.7968
EE	0.0617	0.2930	0.2524	0.2524	0.0774	0.1154	0.0824	0.1052	0.8802
EL	0.0547	0.0995	0.2118	0.2118	0.5270	0.0482	0.2274	0.0487	0.8361
ES	0.1375	0.1141	0.1921	0.1921	0.7482	0.0705	0.0286	0.0585	0.8197
FI	0.0552	0.1759	0.1852	0.1852	0.0810	0.2314	0.0429	0.1552	0.8203
FR	0.2255	0.0557	0.1571	0.1571	0.4648	0.1028	0.0615	0.1082	0.6410
HR	0.2279	0.1049	0.2181	0.2181	0.5979	0.1197	0.0234	0.0681	0.9079
HU	0.1375	0.1979	0.2846	0.2846	0.2457	0.1663	0.0635	0.1194	0.8397
IE	0.1565	0.0660	0.1504	0.1504	0.7040	0.0541	0.0181	0.0586	0.7332
IS	0.0392	0.0913	0.2253	0.2253	0.3077	0.1237	0.0587	0.1785	0.8971
IT	0.1301	0.1154	0.2504	0.2504	0.6966	0.0601	0.0371	0.0511	0.6947
LT	0.2183	0.1911	0.2503	0.2503	0.1229	0.2904	0.0712	0.1111	0.7050
LU	0.0393	0.1860	0.2329	0.2329	0.5786	0.1048	0.0551	0.0602	0.7410
LV	0.1014	0.2670	0.2443	0.2443	0.0884	0.2197	0.0880	0.1184	0.4730
MT	0.1059	0.1003	0.2471	0.2471	0.9210	0.0096	0.0018	0.0181	0.5777
NL	0.0326	0.0840	0.2078	0.2078	0.6687	0.0580	0.0176	0.0894	0.5943
NO	0.0287	0.1001	0.2289	0.2289	0.2748	0.0881	0.0546	0.2055	0.9322
PL	0.0764	0.1603	0.2579	0.2579	0.2160	0.1151	0.2778	0.0976	0.6551
PT	0.0782	0.1166	0.2681	0.2681	0.4434	0.1420	0.1584	0.0748	0.5472
RO	0.1139	0.1346	0.2753	0.2753	0.3544	0.0901	0.2524	0.0663	0.8593
SE	0.0235	0.1300	0.2628	0.2628	0.2782	0.0629	0.0354	0.2588	0.7759
SI	0.1770	0.0834	0.2660	0.2660	0.3358	0.1958	0.0630	0.0935	0.7480
SK	0.1342	0.2196	0.2969	0.2969	0.146	0.2058	0.0362	0.1608	0.706
UK	0.0817	0.1337	0.2369	0.2369	0.3674	0.1279	0.0049	0.154	0.6622

Source: EU-SILC, 2011. Averages are calculated using sampling weights; Only the four most frequent answers for father and mother occupation are reported; Tenancy is the share of respondents reporting living in a house owned by the family when they were around 14.

Table 3A: Missing circumstances

				Father				Mother				
	Sex	Birth area	Tenancy	Birth area	Citizen	Occupation	Education	Birth area	Citizen	Occupation	Education	average share of missing
AT	0%	0%	2%	1%	1%	2%	3%	1%	1%	1%	2%	1%
BE	0%	0%	4%	3%	2%	9%	7%	4%	3%	17%	7%	6%
BG	0%	0%	6%	4%	4%	13%	7%	5%	4%	13%	6%	7%
CH	0%	0%	12%	11%	9%	11%	15%	11%	8%	9%	13%	11%
CY	0%	0%	1%	0%	0%	1%	1%	0%	0%	1%	1%	1%
CZ	0%	0%	24%	25%	26%	25%	27%	24%	24%	25%	24%	25%
DE	0%	0%	5%	4%	4%	12%	16%	3%	3%	9%	16%	8%
DK	0%	0%	53%	52%	53%	61%	59%	52%	52%	58%	55%	55%
EE	0%	0%	3%	2%	4%	6%	10%	3%	5%	5%	6%	5%
EL	0%	0%	4%	1%	1%	5%	17%	1%	1%	3%	19%	6%
ES	0%	0%	2%	2%	1%	4%	5%	2%	1%	2%	4%	3%
FI	0%	0%	52%	50%	50%	67%	51%	51%	50%	66%	50%	54%
FR	0%	0%	3%	3%	3%	6%	9%	2%	2%	3%	7%	4%
HR	0%	0%	9%	6%	6%	12%	17%	6%	6%	8%	14%	9%
HU	0%	0%	3%	2%	2%	6%	4%	1%	1%	4%	2%	3%
IE	0%	0%	26%	25%	25%	27%	28%	25%	25%	26%	27%	26%
IS	0%	0%	56%	56%	56%	57%	59%	56%	56%	57%	58%	57%
IT	0%	0%	2%	0%	0%	1%	0%	1%	0%	0%	0%	1%
LT	0%	0%	5%	5%	6%	13%	11%	4%	6%	9%	7%	7%
LU	0%	0%	2%	0%	0%	2%	2%	1%	0%	1%	2%	1%
LV	0%	0%	2%	2%	8%	4%	6%	2%	9%	4%	4%	4%
MT	0%	0%	7%	6%	6%	12%	9%	6%	6%	7%	8%	7%
NL	0%	0%	50%	50%	51%	51%	54%	50%	50%	50%	53%	51%
NO	0%	0%	50%	52%	52%	51%	51%	49%	49%	50%	50%	51%
PL	0%	0%	11%	7%	7%	13%	11%	7%	7%	11%	9%	9%
PT	0%	0%	2%	1%	1%	1%	4%	1%	1%	1%	3%	2%
RO	0%	0%	7%	0%	0%	22%	9%	5%	0%	13%	7%	7%
SE	0%	0%	54%	55%	54%	90%	70%	55%	54%	84%	64%	65%
SI	0%	0%	64%	64%	100%	64%	65%	64%	100%	64%	64%	72%
SK	0%	0%	4%	1%	1%	6%	2%	1%	1%	4%	2%	3%
UK	0%	0%	17%	16%	16%	20%	21%	15%	16%	17%	20%	18%

Source: EU-SILC, 2011.

Table 4A: Dominance Relations

	AT	BE	BG	CH	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IS	IT	LT	LU	LV	MT	NL	NO	PL	PT	RO	SE	SI	SK	UK
AT	.																														
BE		.																													
BG	>	>	.																												
CH	<*	<*	<*	.																											
CY	<*	<*	<**	<	.																										
CZ	>	>		>**	>**	.																									
DE	<	<	<		>	<*	.																								
DK			<	>	>	<	>	.																							
EE	>**	>***	>	>***	>***	>	>***	>*	.																						
EL	<**	<**	<**	<		<**	<	<*	<***	.																					
ES	<***	<***	<***	<	<	<***	<	<**	<***		.																				
FI	<*	<	<*			<**		<	<***			.																			
FR	<**	<	<**		>	<**		<	<***	>	>	>	.																		
HR	>***	>***	>***	>***	>***	>***	>***	>***	>	>***	>***	>***	>***	.																	
HU	>***	>***	>**	>***	>***	>**	>***	>***		>***	>***	>***	>***	<*	.																
IE	<	<	<***			<**		<	<***					<***	<***	.															
IS		<	<		>	<		<	<*	>	>*	>	>	<***	<***	>	.														
IT	<***	<	<***			<**		<	<***					<***	<***			.													
LT	>*	>**	>	>***	>***	>	>***	>		>***	>***	>***	>***	<*	<	>**	>**	>***	.												
LU		>*		>***	>***		>***		<	>***	>***	>***	>***	<***	<***	>***	>	>***	<	.											
LV	>	>	>	>***	>***	>	>**	>	<	>***	>***	>**	>***	<**	<	>**	>	>***	<	>	.										
MT	<***	<**	<***	<	<	<***	<	<*	<***	<		<	<	<***	<***		<	<	<***	<***	<***	.									
NL	<**	<*	<**			<**		<*	<***					<***	<***		<		<***	<***	<***	>	.								
NO	<		<	>	>	<	>	<	<*	>	>*	>	>	<***	<***	>		>	<*	<	<	>*	>	.							
PL	>	>		>**	>**		>*	>	<	>**	>***	>**	>**	<***	<*	>***	>	>***	<		<	>***	>**	>	.						
PT	>***	>***	>***	>***	>***	>***	>***	>***		>***	>***	>***	>***	<		>***	>***	>***		>**		>***	>***	>***	>***	.					
RO	<	<	<	>	>	<	>	<	<	>	>*	>	>	<*	<**	>		>	<	<	<	>	>		<	<	.				
SE					>				<*	>*	>**	>	>	<***	<***	>			<*	<	<	>*	>		<	<***		.			
SI	>*	>**	>	>***	>***	>	>***	>	<	>***	>***	>***	>***	<**	<	>***	>*	>***		>	>	>***	>***	>*	>		>	>*	.		
SK	>	>*	>	>**	>***	>	>***	>	<	>***	>***	>**	>***	<***	<	>***	>*	>***	<			>***	>***	>*	>		>	>*		.	
UK	<	<	<**			<*		<	<***		>			<***	<***	>			<**	<**	<*	>			<*	<***			<**	<***	.

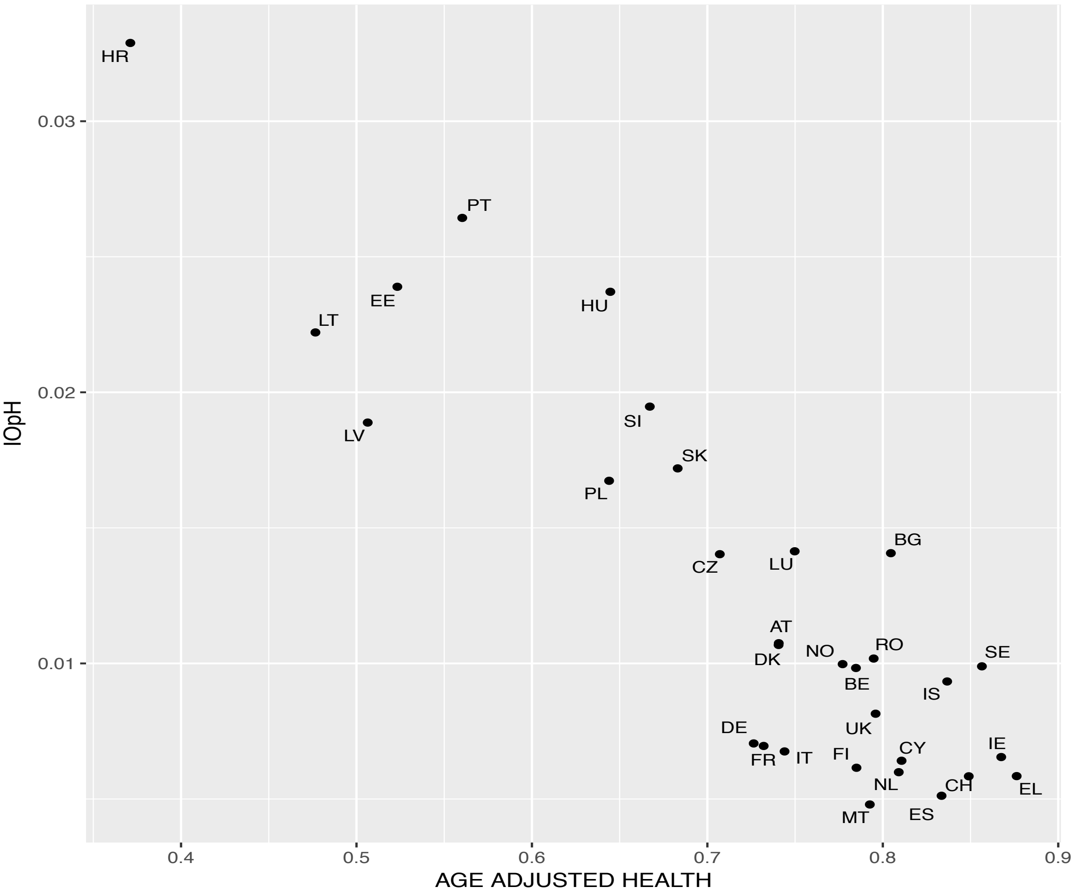
*Source: EU-SILC, 2011. Dominance is verified for each possible number of latent classes based on 200 bootstrap replications. Statistical significance: ***=0.01, **=0.05, *=0.1. No star means that the two curves do not cross, so there is dominance if you only look at point estimates, but the difference is never statistically significant.*

Table 5A: Health and inequality indicators

	IOpH (1)			SAH	Private health expenditure	Total inequality	IOp
Country	mean	low	high	Age-adj. (1)	% GDP (2)	Gini (3)	Income (4)
AT	0,01	0,007	0,013	0,7407	2,72	27,4	0,088
BE	0,009	0,006	0,013	0,7847	2,42	26,3	0,091
BG	0,013	0,01	0,015	0,8046	3,12	35,0	0,134
CH	0,005	0,004	0,007	0,8490	3,91	29,7	0,09
CY	0,006	0,004	0,008	0,8107	4,00	29,2	0,08
CZ	0,013	0,01	0,016	0,7071	1,19	25,2	0,051
DE	0,006	0,004	0,009	0,7264	2,63	29,0	0,079
DK	0,01	0,006	0,014	0,7406	1,60	26,6	0,02
EE	0,022	0,017	0,028	0,5233	1,12	31,9	0,101
EL	0,005	0,003	0,008	0,8763	3,07	33,5	0,109
ES	0,005	0,004	0,006	0,8335	2,48	34,0	0,12
FI	0,006	0,003	0,009	0,7850	2,25	25,8	0,028
FR	0,006	0,005	0,008	0,7321	2,46	30,8	0,098
HR	0,03	0,023	0,036	0,3711	1,54	31,2	0,076
HU	0,022	0,019	0,024	0,6447	2,81	26,9	0,108
IE	0,006	0,004	0,009	0,8675	2,62	29,8	0,078
IT	0,009	0,005	0,013	0,8367	2,31	32,5	0,097
IS	0,006	0,005	0,008	0,7440	1,68	23,6	0,016
LT	0,02	0,016	0,024	0,4766	1,90	33,0	0,067
LU	0,013	0,01	0,016	0,7498	1,08	27,2	0,136
LV	0,017	0,013	0,022	0,5064	2,08	35,1	0,111
MT	0,004	0,002	0,007	0,7926	3,14	27,2	0,072
NL	0,005	0,003	0,008	0,8091	1,44	25,8	0,019
NO	0,009	0,005	0,014	0,7771	1,42	22,9	0,023
PL	0,015	0,013	0,018	0,6440	1,96	31,1	0,099
PT	0,024	0,02	0,029	0,5603	3,37	34,2	0,127
RO	0,009	0,006	0,024	0,7948	1,15	33,5	0,111
SE	0,009	0,004	0,016	0,8565	1,78	24,4	0,031
SI	0,018	0,014	0,022	0,6672	2,40	23,8	0,036
SK	0,016	0,013	0,019	0,6831	2,40	25,7	0,046
UK	0,007	0,005	0,01	0,7959	1,57	33,0	0,079

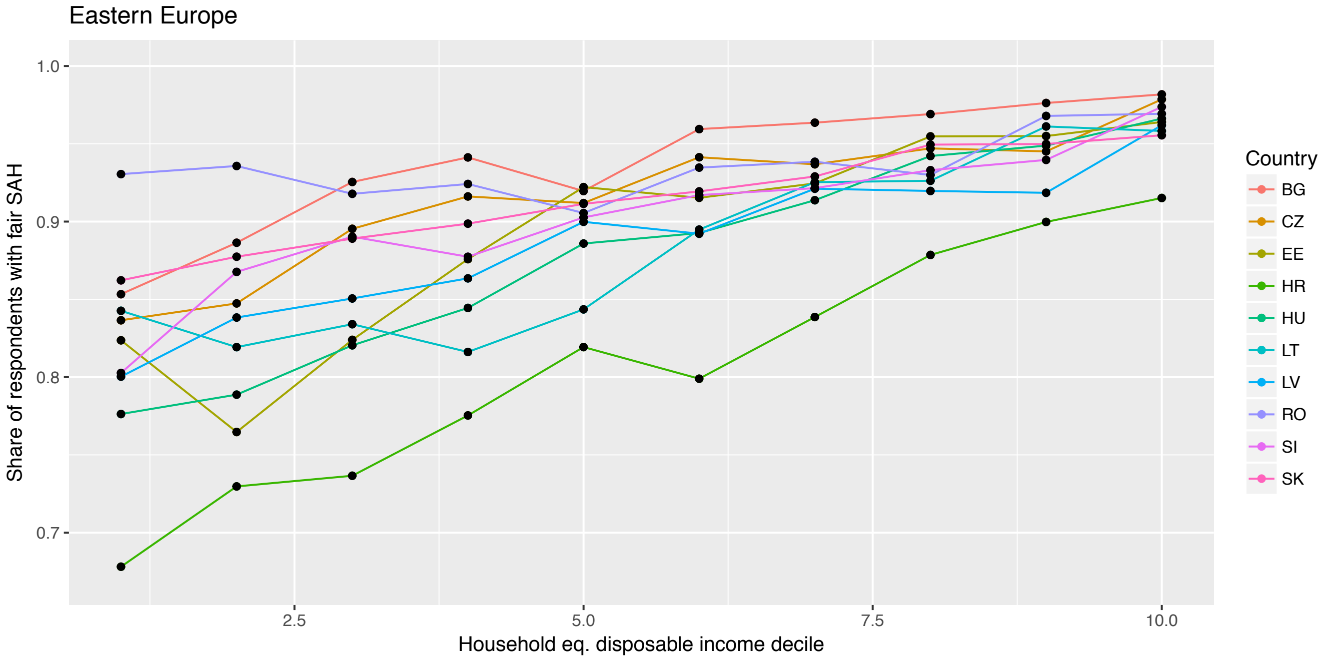
Source: (1) EU-SILC, 2011; (2) World Development Indicators (World Bank); (3) Eurostat; (4) Inequality of opportunity in income as estimated by Brunori et al. (2019).

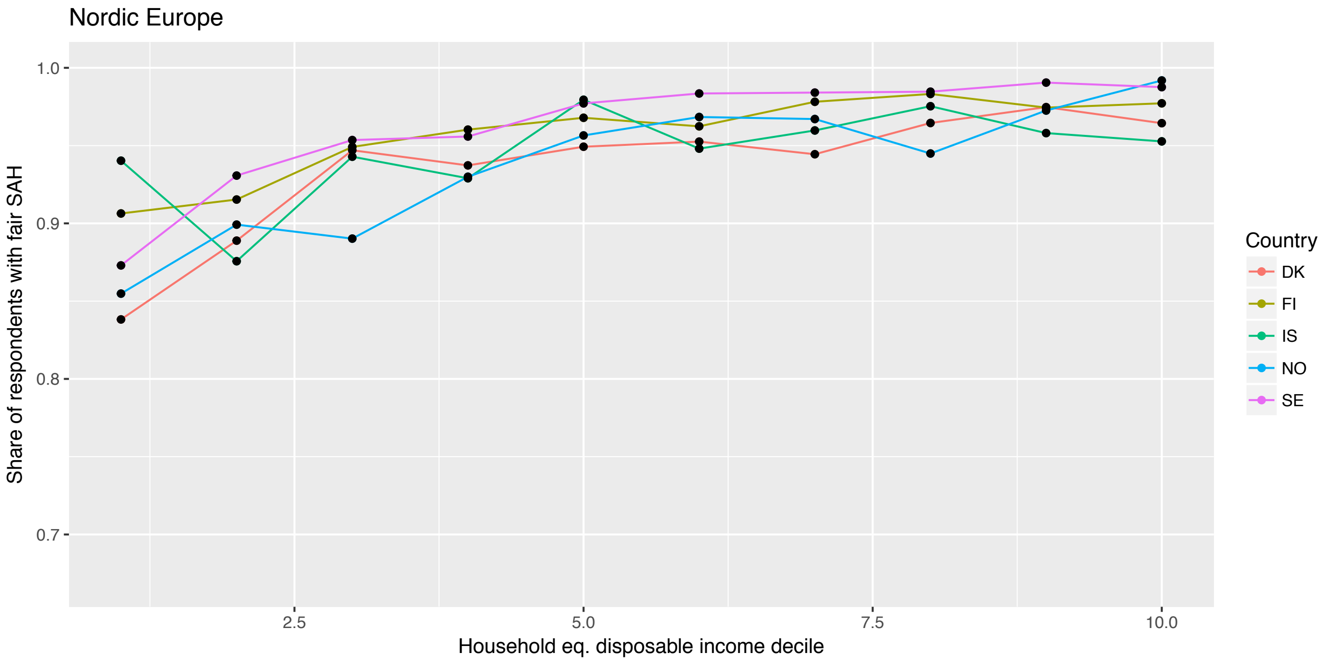
Figure 6A: Correlation between IOpH and Age-adjusted health

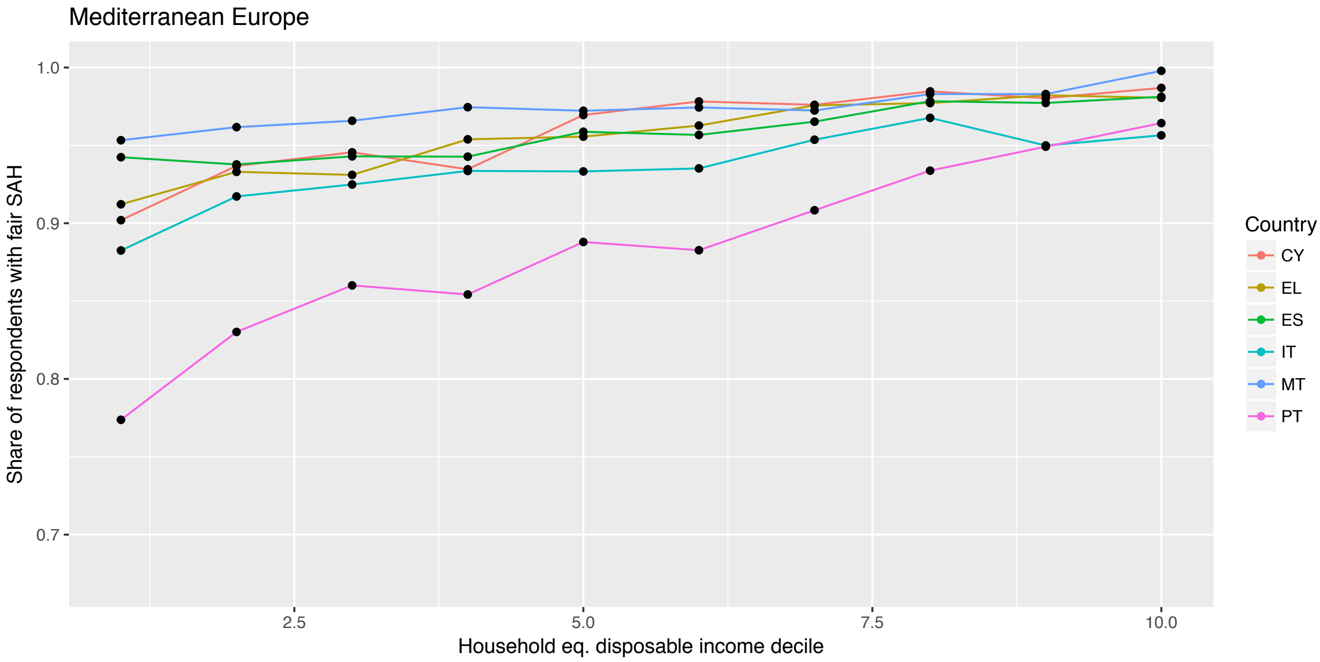


Source: see Table 5 A

Figure 7A: Income-SAH trend across European regions






Mediterranean Europe
Share of respondents with fair SAH
Household eq. disposable income decile
1.0
0.9
0.8
0.7
2.5
5.0
7.5
10.0
Country
CY
EL
ES
IT
MT
PT

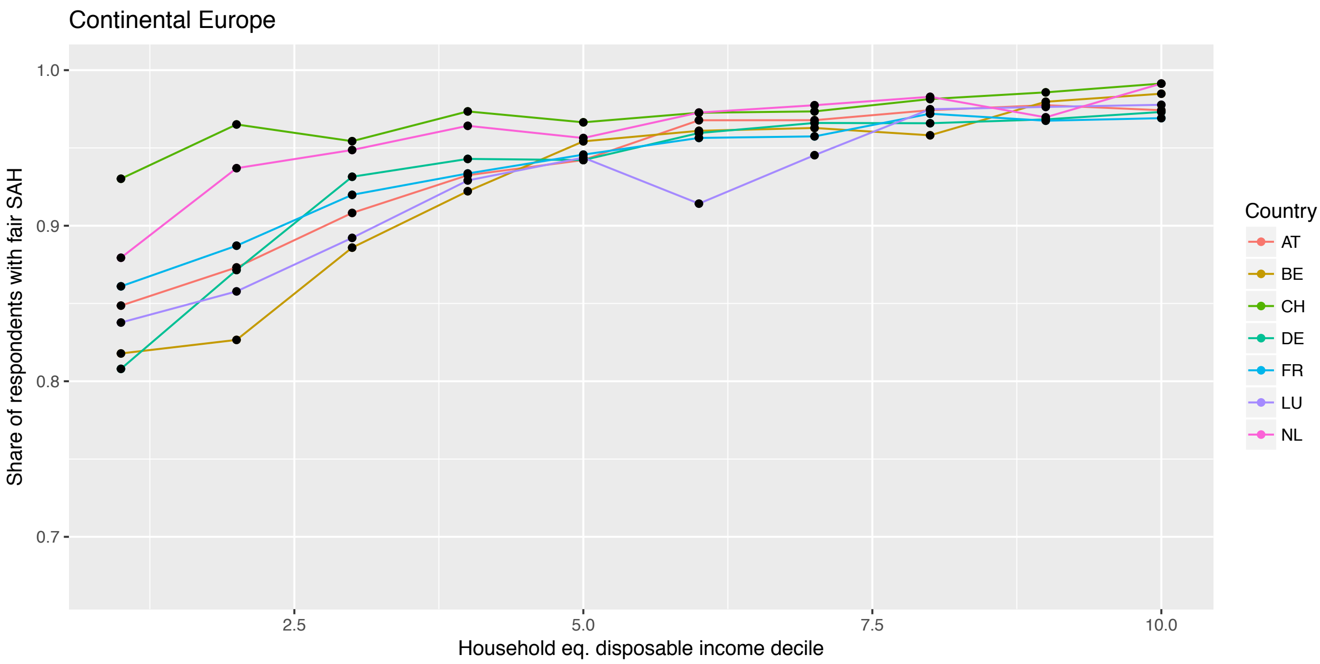
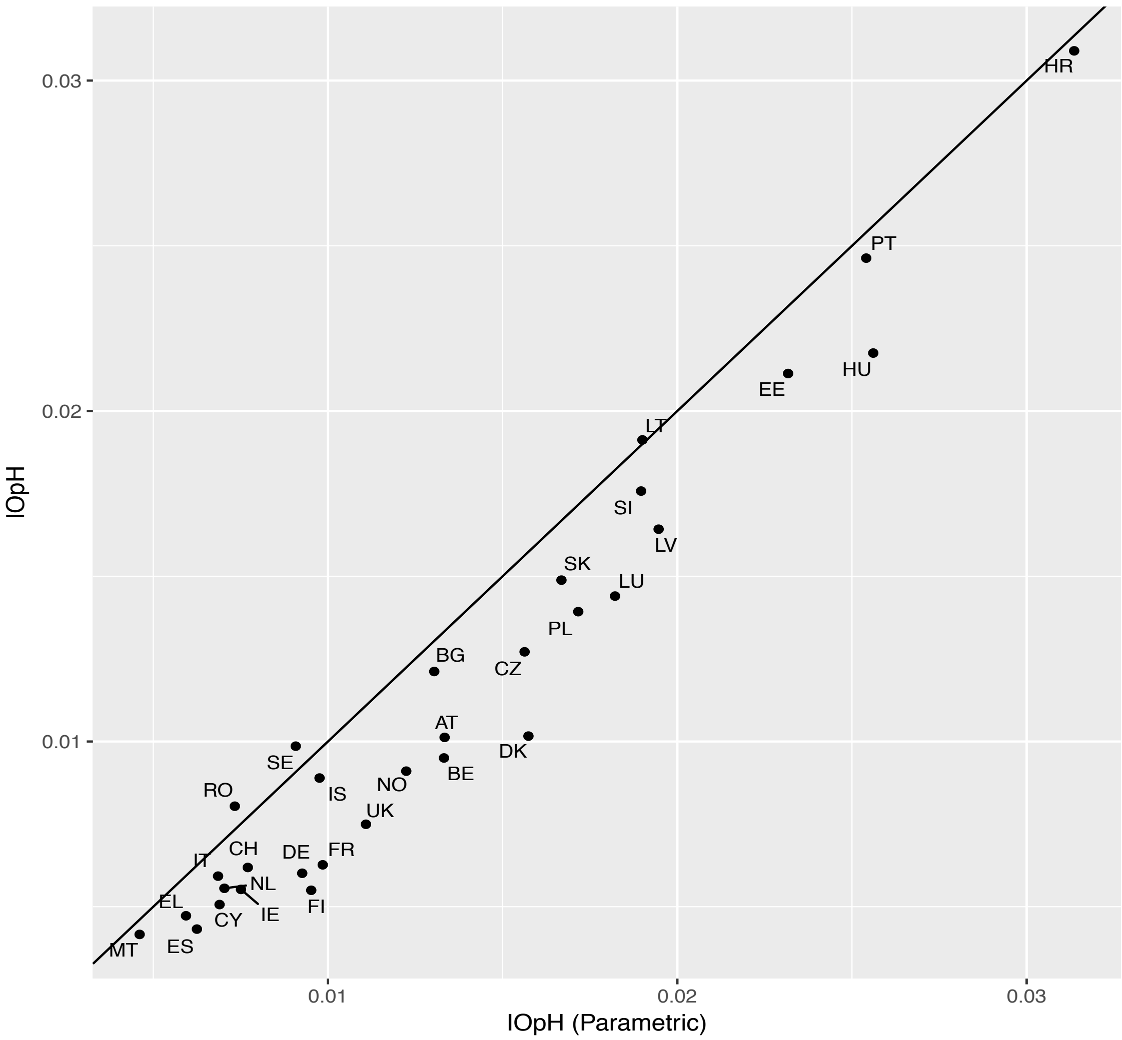

Continental Europe
Share of respondents with fair SAH
Household eq. disposable income decile
1.0
0.9
0.8
0.7
2.5
5.0
7.5
10.0
Country
AT
BE
CH
DE
FR
LU
NL

Figure 8A: Comparison of ranking of European countries according to IOpH latent types vs. parametric method



Online appendix

Brunori, P. Guidi C. F., Trannoy A. "Ranking populations in terms of Inequality of health opportunity: A flexible latent type approach".

The estimation of latent types

Let assume that types are categories of a latent variable by definition unobservable. What we may observe are indicators, unidimensional manifestations of types' membership. For example, we can observe whether an individual experienced poverty in childhood, this is an indicator of a higher probability to belong to more disadvantaged latent types. The identification of latent types is made possible using latent class analysis. Here, following Collins and Lanza (2010), we discuss how the formal link connecting indicators (observable circumstances) and latent classes (types) can be estimated.

Assume we have $c = 1, \dots, C$ observable circumstances, each circumstance can take $r_c = 1, \dots, R_c$ categories (continuous circumstances are excluded). The partition obtained interacting all possible categories of all observable circumstances is made of $N = \Pi_{c=1}^{C} R_c$ types. Each type is described by a list of characteristics, a complete response pattern, $y = (r_1, \dots, r_C) \in Y$. Where Y represents all possible response patterns, that is all possible combinations of categories one for each circumstance. Each response pattern has a probability $P(Y = y)$ and $\sum_{n=1}^{N} P(Y = y) = 1$. Λ represents the categorical latent variable made of $l = 1, \dots, L$ latent classes. $\lambda_l$represents the prevalence of class l and $\sum_{l=1}^{L} \lambda_l = 1$.

The probability $\rho_{c,r_c|l}$ is the probability of being characterized by value r_c for circumstance c conditional on membership of latent class l. The set of parameters ρ represents the relationship between each circumstance and each latent class. The vector of item response probabilities for a particular circumstance conditional on membership to a particular class sum to 1:
$\sum_{r_c=1}^{R_c} \rho_{c,r_c|l} = 1 \quad \forall\, c = 1, \dots, C;\ \forall\, l = 1, \dots, L.$

Finally, let y_c be the element c of pattern response y. Let define an indicator function $I(y_c = r_c)$ which takes value 1 if the value $c = r_c$ and takes value 0 otherwise.

The fundamental assumption of latent class analysis is that, conditional on belonging to a particular latent type, all observable circumstances are independent. Then the link between being characterized by a particular pattern of circumstances and belonging to a particular latent type is:

$$P(Y = y) = \sum_{l=1}^{L} \lambda_l \Pi_{c=1}^{C} \Pi_{r_c=1}^{R_c} \rho_{c,r_c|l}^{I(y_c = r_c)} \qquad \text{(A.1)}$$

Where $P(Y = y \cap \Lambda = l) = \Pi_{c=1}^{C} \Pi_{r_c=1}^{R_c} \rho_{c,r_c|l}^{I(y_c = r_c)}$ is the probability of observing a particular circumstance pattern *y* conditional on membership in latent class *l*.

The posterior probability that individual *i* belongs to latent class *l*, conditional on the observed values of the observable circumstances *y*, can be obtained applying the Bayes' theorem:

$$P(\Lambda = l | Y = y) = \frac{P(Y = y | \Lambda = l) P(\Lambda = l)}{P(Y = y)} \qquad \text{(A.2)}$$

Substituting eq. (A.1) into eq. (A.2) and recalling that $\lambda_l = P(\Lambda = l)$ we write such probability as:

$$P(Y = y) = \frac{\lambda_l\left(\Pi_{c=1}^{C}\Pi_{r_c=1}^{R_c}\ \rho_{c,r_c|l}^{I(y_c=r_c)}\right)}{\sum_{l=1}^{L}\lambda_l\Pi_{c=1}^{C}\Pi_{r_c=1}^{R_c}\ \rho_{c,r_c|l}^{I(y_c=r_c)}} \quad \text{(A.3)}$$

Equation A.3 estimate the probability of an individual characterized by a certain pattern of response $(Y = y)$ to belong to the latent type *l*.

It is possible to obtain and estimation of the parameters ρs and λs by maximum likelihood. The maximum likelihood is estimated using the expectation-maximization algorithm. This log-likelihood function is identical in form to the standard finite mixture model log-likelihood (Linzer and Lewis, 2011). However, it is important to notice that the number of parameters to be estimated grows rapidly with the number of circumstances considered (*C*), the number of values each circumstance can take (R_c), and the number of latent classes L. The exact number of parameters is: $\aleph = L\sum_{c=1}^{C}(R_c - 1)(L - 1)$.

Once the posterior probabilities to belong to each latent class are estimated the highest estimated probability is used to associate each individual in the sample to one and only one latent type.

References

Collins Linda and Stephanie Lanza (2010), "*Latent Class and Latent Transition Analysis*", John Wiley and Sons. Hoboken, New Jersey.

Linzer D, Lewis J. (2011), "poLCA: An R package for polytomous variable latent class analysis", Journal of Statistical Software, Vol. 42 (10) pp. 1-29.