HEDG Health, Econometrics and Data Group

THE UNIVERSITY of York

HEDG Working Paper 09/24

Measuring income-related inequalities in health using a parametric dependence function

Casey Quinn

July 2009

ISSN 1751-1976

york.ac.uk/res/herc/hedgwp

Measuring income-related inequalities in health using a parametric dependence function

Casey Quinn
University of Nottingham
Nottingham, NG7 2RD

Abstract

Attention has been given recently to the Concentration Index; specifically, corrected versions have been generated that supersede the original with properties such as transform invariance, reversal invariance and transfer invariance. While previous studies have promoted a transformed or normalised index to overcome these problems, I propose, in this paper, two novel approaches to a direct parametric model for dependence as a measure of inequality in the distributions of health and income. These are the copula and quantile regression using jackknifed samples.

As well as accommodating any form of health or income, and being robust to invariance criteria, both methods parameterise the measure of inequality directly, rather than indirectly through functions on one of the marginals. Results from an illustrating example using the Survey of Health, Retirement and Ageing in Europe suggest that such inequality in these countries is not explained well by covariates on age, gender, education and lifestyles.

JEL classification: C46, C51, I10
Keywords: Health Inequality, Non-Continuous data, Copulas, Quantile Regression

1 Introduction

In a recent exchange, Erreygers (2009a, 2009b) and Wagstaff (2009) debated the merits of 'their' respective indices of socioeconomic related inequalities in health (which, in this study, translates to income-related inequalities in Self-Assessed Health; however, the results here apply more broadly). Erreygers' (2009a) argument took the view that four consistent properties were required of a measure of income-related inequalities in Self-Assessed Health (SAH): transform invariance, mean invariance, transfer invariance and order invariance.

In both the Erreygers index E and the Wagstaff index W (Erreygers 2009a), the Concentration Index was re-scaled so that it overcame particular weaknesses: principally, the failure to be bounded properly $CI \in (-1, 1)$. In this paper, I instead propose a different paradigm of measurement: direct parameterisation of dependence using demographic, socioeconomic and lifestyle covariates. As well as rank-correlation meeting the criteria enumerated in Erreygers (2009a, 2009b), assigning covariates directly can offer more precise information for policy responses.

Both methods define socioeconomic inequalities in health as the dependence between the distributions of Self-Assessed Health and Income, in concordance with the principles of inequality found in, for example, Wagstaff, *et al.* (1991), Bommier and Stecklov (2002) and the International Society for Equity in Health (ISEqH, 2001).

The first method employed uses copulas to bind the distributions of the margins of health and socioeconomic status. Copulas are rank-based measures of association between monotonic transformations of random variables, rather than the variables themselves. In essence, a copula is a multivariate distribution functions whose margins consist of univariate uniformly-distributed variables. Like the concentration index, copulas can jointly estimate health and income. Because dependence is tractably separate from the margins in the bivariate distribution function, however, copulas can be constructed such that dependence itself is the response variable. Unlike other methodological approaches to measuring inequality, this gives the researcher direct access to the factors determining socioeconomic inequalities in health.

The second method uses jackknifing techniques to generate resampled data containing sample dependence and sample means or proportions of the covariates. Quantile regression methods can then be used to estimate, at the country level, the effect of the population-level characteristics on the dependence between health and income in the population. As well as generating policy-relevant interpretations of the results, this method overcomes the problems associated with using dependence, a population parameter that is not indexed by the individual.

I present here results for concentration indices as well as copulas and quantile regression, for an international comparison of income-related inequalities in SAH using data on retirees in 11 European countries. The results show that parameterisation of dependence generates different information on size and statistical significance of the factors of inequality when measured directly, rather than indirectly through Generalised Linear Models for health.

2 The Concentration Index

There are several representations of the Concentration Index; the most useful for this analysis is the "convenient covariance" representation, in which each individual's health h_i (from a distribution of health H with mean μ_h) is indexed against their rank R_i in the distribution of income Y (Kakwani 1980; Wagstaff *et al.* 2003). Thus

$$\begin{aligned} CI &= \frac{2}{\mu_h}\sum_{i=1}^{n}(h_i-\mu_h)(R_i-\frac{1}{2}) \\ &= \frac{2}{\mu_h}\text{cov}(h,R) \end{aligned} \tag{1}$$

The Concentration Index *CI* is the scaled covariance between the health of the individual and their rank in the income distribution: $CI = 0$ when there is no inequality - i.e. no observed association between income and SAH - and $-1 \leq CI \leq 1$ due to the $\frac{2}{\mu_h}$ term. Covariance forms the basis of the measurement of socioeconomic-related inequalities in health, according to the criteria of Wagstaff *et al.* (1991) and Bommier and Stecklov (2002), among others.

Concentration indices are appropriate for ratio-scale random variables but not ordinal SAH (Wagstaff *et al.* 1991, 2005; Erreygers (2006, 2009a); O'Donnell *et al.* 2008). Allison and Foster (2004) demonstrated the non-robustness of the mean as a reference point for measuring inequality of a ordinally-scaled variables because transformations (of the mean, linearly, or of the distribution non-linearly) will place the mean between different categories in the ordinal scale. For example, for five people whose SAH is distributed uniformly amongst the categories $S_1 = [1, 2, 3, 4, 5]$, $\mu_{h_1} = 3$. Under linear transformations, μ_h remains the midpoint, however non-linear transformations generate skew, moving the mean. The scales $S_2 = S_1^2$ and $S_3 = e^{S_1}$ move the mean above the third category; $S_4 = \text{Ln}(S_1)$ move it to below the third category. In fact, the greater the value placed upon the upper categories, the greater the increase in the mean, and vice-versa.

This means that any mean-scaled measure of inequality will not be stable under re-scaling - particularly when, in practice, the distribution of individual SAH will almost certainly not be uniform. This is relevant to measuring inequality with SAH because the cardinal scale applied to represent the 'true' value of health will vary in this manner: in particular, whether low or high values of SAH are relatively more important (to avoid or to achieve, respectively) is a normative problem. Any given participant in such a comparison might - rightly - complain that such a comparison unfairly represented them due to the cardinal scale assumed for SAH.

2.1 The Erreygers and Wagstaff indices

See Erreygers (2009a, 2009b) and Wagstaff (2005, 2009) for a comprehensive presentation and discussion of these. The former paper presents the iterative 'corrections' for the Concentration Index. Erreygers (2009a) proposed what he called a rank-dependent family of indicators:

$$I(h) = 2f(h)\sum_i \frac{h_i}{n+1-2R_i} \tag{2}$$

He showed that the Concentration Index, normalised Concentration Index of Wagstaff (2005) and Generalised Concentration Index of Wagstaff *et al.* (1991) all were specific forms of his general family, the differences lying in the form of $f(H)$, some function $f(H) > 0$ capturing health and/or population characteristics. By allowing H to contain, freely, μ_h, b_h, a_h, etc., he is able to make a general class containing the others.

His 'corrected' Concentration Index is the specific form:

$$E(h) = \frac{16}{n^2(b_h - a_h)}\sum_i \frac{h_i}{n+1-2R_i} \tag{3}$$

A more straightforward solution however is leave the Concentration Index itself unchanged - including separation of the margins of health and income - and re-configure SAH itself.

2.2 Non-parametric extension to continuity of SAH

Discrete SAH can be transformed using *continuous extension*: the discrete random variable h is said to be continued by U when, for $m = 1, .., M$ levels of an ordinal random variable

h, $h^*_{(m)} = h_{(m)} + (h_{(m)} - h_{(m-1)})(U-1)$, where $U \subset (0,1)$ has a strictly increasing CDF independent of h (see, for example, Denuit and Lambert 2005, Mesfioui and Tajar 2005).

This approach preserves concordance within the distribution of health h and deals with the problem of non-continuity, however it cannot preserve rank which is now one of only five values in the distribution of SAH.[1] Dependence will be weaker as a result of the continuation, and will also vary for such transformations, making results non-replicable by different analysts; however, because the mapping of $h^* \rightarrow h$ is invariant to this (provided $U \subset (0,1)$ in this case) the results in terms of correlation will not vary. More importantly, continuation of a random variable will preserve the mean and support; because of this, the effects of mean-scaling the Concentration Index will also be preserved.

A similar approach can be taken to extending rank by continuation when, with discrete data, ranks are tied. For ordinal data with ranks $R_{m \in 1,..,M}$, continued rank $R^*_{(m)} = R_m + (R_m - R_{m-1})(U-1)$ where $U \subset (0,1)$ has a strictly increasing CDF independent of R.[2] Unlike continuation by extension of SAH itself, that of the rank of SAH will preserve the midpoint of the distribution as the mean - in this case 0.5, such that Equation (4) holds.

2.3 The Rank-Concentration Index

In Quinn (2009), I proposed an alternative to correcting the Concentration Index: correcting, instead, health. This replaced ordinal SAH with its rank, or empirical CDF:

$$F(h) = \frac{\#[i : 1 \leq i \leq n, h_i \leq h]}{n} \tag{4}$$

So that the Concentration Index simplifies dramatically. Cardinal scales S_1 to S_4 will contain the same cut-points in relative terms, allowing for the proportions of each latent scale, but will generate a different mean. However, any rank-preserving transformation will generate the same rank-order amongst observations. Scaling Equation (1) by the mean of the rank reduces it to the covariance of the ranks.

2.4 Decomposing rank-covariance

Decomposition is undertaken to isolate (i) inequality due to a given source (in this case of health); (ii) the reduction in inequality that would result if a given factor of health ceased to be a factor, or if it was distributed evenly across all ranks of income; and (iii) the percentage of inequality that would be observed if all other factors of health were distributed evenly across ranks of income.[3] It functions much like a sub-group analysis of the inequality in the distribution of income, grouped by the factors explaining health.

The index in Equation (4) can be decomposed according to the same principles first identified by Wagstaff *et al.* (2003), in their analysis of socioeconomic inequalities in height-for-age in Vietnam. They showed that, for a linearly-additive regression model for health h, (O'Donnell *et al.* 2008)

$$h = \alpha + \sum_j \beta_j x_j + \varepsilon \tag{5}$$

for $J = 1, .., J$ regressors, the decomposed concentration index is such that

$$CI_h = \sum_j CI_j \frac{\hat{\beta}_j \bar{x}_j}{\mu_h} + \frac{CI_\varepsilon}{\mu_h} \tag{6}$$

where CI_j is the concentration index for each x_j as defined in Equation(1) and CI_ε is the generalised concentration index for error ε, given also by the residual of the concentration indices.

For a well-specified model $CI_\varepsilon = 0$ and the concentration index CI_h is the weighted sum of decomposed concentration indices of the factors of health only. Thus the decomposed concentration index, originally the covariance between health and income, is made up instead of the linearly-additive explanators of health and income, variable-by-variable.

In this instance Generalised Linear Models (GLMs) can be used for SAH, such that

$$G(h) = \alpha + \sum_j \beta_j x_j + \varepsilon \tag{7}$$

where $G(h)$ can be an ordered or binary choice model, normal or logistic regression, etc. Decomposition follows Equation (6). Cardinal scales S_1 to S_4 will contain the same cut-points in relative terms, and any CDF $G(H)$ will generate the results in Equation (4). GLM can be applied to SAH continued by extension, also: all such extensions are Lebesgue-measurable. The assumptions of linear covariation between health and socioeconomic status extend to the relationship between the factors of health individually and socioeconomic status. Income is still ranked by R_i, the empirical CDF of income.

The strength of the information gained on socioeconomic inequalities in health, however, can only be as great as the explanatory power in the GLM for SAH used to weight the decomposed Concentration Indices in Equation (6). In a previous analysis (Quinn 2009) little such power existed: in the ordinal probits used, R^2, for example, ranged from 2.5% to nearly 9.5% across the SHARE countries. Explanatory power in the intermediate model affects how well we understand the directly relationship between inequality and the variables of interest. A more appropriate method should factorise the inequality directly.

3 Copulas

A copula is best described as a multivariate distribution function that binds each marginal distribution function to form the joint (Joe 1997); the copula parameterises the dependence between the margins, while the parameters of each marginal distribution function can be estimated separately. For univariate marginal distribution functions $G(h)$ for SAH and $F(y)$ for income, a copula is a function that binds those margins precisely, to form the multivariate distribution function (Smith 2003. See Joe 1997, Nelsen 2006 for examples).

By Sklar's (1959) theorem, one can say that all multivariate distributions have a copula representation, in which each margin is invariant to transformations in every other margin, or independent of the choice of every other marginal distribution. Consider two random variables H, Y with bivariate distribution function $O(h, y) = \Pr(h \leq H, y \leq Y)$ and univariate marginal distributions $G(h)$ and $F(y)$ respectively. Then there exists a copula C that represents the joint distribution function in terms of the margins, such that

$$O(h, y) = C(G(h), F(y); \theta) \tag{8}$$

In particular, copulas are tractable with respect to the margins and the parameter of association (see Smith (2007) for his study of this issue and its application to Fisher Information). Because of this, $G(h)$ and $F(y)$ need not be identical distribution functions, which facilitate more flexible estimation of jointly-dependent variables. However it also means that copulas provide a lot of flexibility over the dependence structure without affecting estimation at the margins. This includes asymmetric and tail-dependence copulas, however it also means that covariates can be used to explain the dependence, rather than treating it as a nuisance parameter.

The dependence parameter θ that binds $G(h)$ and $F(y)$ to form the copula is a measure of rank covariance. That is, the covariance between the ranks, or CDFs, of the random variables. Like the index shown in Equation (4) this means that empirical CDFs can freely be used in place of SAH or income, reducing the problem to one of explaining only dependence itself, without any loss of information.

3.1 Parameterisation of dependence only

Parameterised dependence is a recent development in the use of copulas (Bogaerts and Lessaffre 2008; Nikoloulopoulos and Karlis 2008). A parametric dependence function, rather than a nuisance parameter, facilitates measurement of a direct effect of a covariate on dependence, rather than the indirect approach of decomposition. For this application the dependence parameter will be constructed as a linear combination of covariates - most similar, in fact, to decomposition in the sense of Equations (5) and (6).
Four specifications of copulas are examined: three Archimedean-class copulas (see Nelsen 2006) and the Gaussian copula. First, the Frank, given by (Frank 1979)

$$C(u, v; \theta) = -\frac{1}{\theta} \ln(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}) \tag{9}$$

where $u = G(h)$, $v = F(y)$ and where $\theta \in (-\infty, \infty)\backslash\{0\}$ corresponds to $\tau \in (-1, 1)\backslash\{0\}$.[4]
The second copula is the Clayton, given by (Clayton 1978):

$$C(u, v; \theta) = \max[(u^{-\theta} + v^{-\theta} - 1)^{\frac{-1}{\theta}}, 0] \tag{10}$$

where $\theta \in [-1, \infty)\backslash\{0\}$.

The final Archimedean-class copula is the Gumbel, given by (Nelsen 2006)

$$C(u,v;\theta) = Exp(-((-Ln(u))^{\theta} + (-Ln(v))^{\theta})^{frac1\theta}) \tag{11}$$

where $\theta \in [-1, \infty)$.

Finally the Gaussian copula will be used, given by:

$$C(u,v;\rho) = \Phi_{(1,2)}(\Phi_1^{-1}(u)), \Phi_2^{-1}(v)))) \tag{12}$$

Although the marginal distribution functions in the copula can still incorporate covariates, analysis will be simpler using only Empirical CDFs for both income and health. Both margins are therefore considered as Empirical CDFs following Equation (4). Estimation by canonical maximum likelihood then satisfies (Joe 2005, Yan 2007)

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \ln c(F(h), F(y); \theta) \tag{13}$$

3.2 Goodness of fit

The appropriate copula can be found using information criteria such as the Akaike Information Criterion (AIC), where $AIC = 2k - 2\ln(L)$, for log-likelihood L and k parameters, or Bayesian Information Criterion (BIC), $BIC = k\ln(n) - 2\ln(L)$, where n is the sample size. Since all copula models will be equally parameterised this is the equivalent of selecting the copula with the greatest likelihood. Other approaches can include model-averaging methods described in Bogaerts and Lesaffre (2005). The results presented here will only include the 'best' copula for each country, however the remaining results are available from the author.

4 Jackknifed samples and Quantile Regression

The second approach to parameterising dependence lies in the construction of a resampled dataset using the delete-one jackknife. This approach arises in response to the fact that, unlike a sample mean, measures of concordance or dependence, such as θ in Equations (9) to (12), cannot be indexed by individual. This affects estimation procedures, since there is no individual observation on $\hat{\theta}_i$, nor does θ have a distribution in the population. An $n = n$ pseudo-sample can be constructed from any given sample, containing θ indexed to each jackknifed sub-sample used. This can then be combined with sample means and sample proportions of the variables of interest as covariates.

The implication of this is that, at a the survey level (for example the national level) a model can be used to explain the effect of population levels of the covariates on dependence between SAH and income in the population. For individually-indexed population parameters, such as average income or health, such an approach should generate equivalent estimates whether the original or the pseudo-sample was used, assuming exchangeable, homoskedastic observations in the original sample (Good 2002). A similar motivation underlies the use of jackknifing for cross-validation, for example (Shao and Tu 1995).

4.1 Quantile Regression

Once a pseudo-sample has been constructed, it will contain $\hat{\theta}_i$, $i \in (1, .., n)$ corresponding to each sub-sample used during jackknifing. Then, $\hat{\theta}$ can be defined according to its conditional qth quantile function, thus:

$$Q_\theta(q|x) = \alpha + \sum_i \beta_q x_i \tag{14}$$

which is then solved with $\hat{\beta}_q$, where

$$\hat{\beta} = \underset{\beta \in B}{\arg\min} \sum_{i=1}^{n} \rho_q(\theta_i - \beta x_i) \tag{15}$$

where ρ_q is some loss function defined according to Koenker (2005).

Quantile regression is particularly appropriate to this problem because of the uniqueness of dependence captured by a given copula, in particular the domain of θ. For example, the coefficients from the Frank copula, where $\theta \in (-\infty, \infty)\backslash\{0\}$, are not directly comparable with those of the Gaussian copula, in which $\theta \in (-1, 1)$.

Like the procedure regarding copulas with parameterised dependence, the optimal model will be selected using information criteria. The estimate $\hat{\theta}_i$ for $i \in (1,..,n)$ replications will be retained in the pseudo-sample for quantile regression estimation.[5]

5 Illustrating example: the Survey of Health, Ageing and Retirement in Europe

The Survey of Health, Ageing and Retirement in Europe (SHARE) is ... "a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of more than 30,000 individuals aged 50 or over."[6] Eleven European countries are included in the current data: Denmark, Sweden, Austria, France, Germany, Switzerland, Belgium, the Netherlands, Spain, Italy and Greece. Although Israel also contributed to baseline data it is not included in this analysis.[7] This paper uses individuals over 40 years of age and Wave 1 (2004) of the SHARE data. It is therefore only cross-sectional, with each country separately identified, although the methods here can be extended to the SHARE panel as it expands.

The SHARE data contain information on a range of social and economic characteristics: physical health and functioning, including lifestyles and health care service utilisation; psychological health and functioning, labour-market activity including work during retirement; socioeconomic variables including income transfers and consumption; housing; education and social support. Preliminary descriptive analysis of the data has been published by the Mannheim Research Institute for the Economics of Aging (Börsch-Supan *et al.* 2005).[8][9]

The key variables in this paper are SAH and Gross Equivalised Household Income, PPP-adjusted. Gross PPP-adjusted Household Income can be found in a supplementary set of the SHARE data, labelled 'imputations': individual equivalents were taken using household size. Figures 1 and 2 show these variables descriptively for each country.

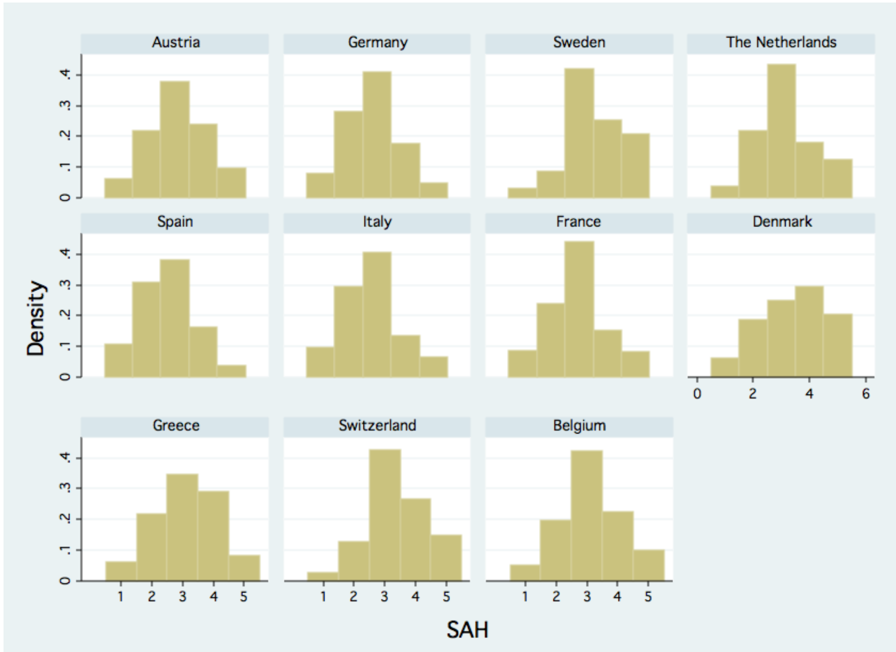

Figure 1: SHARE Self-Assessed Health by Country.

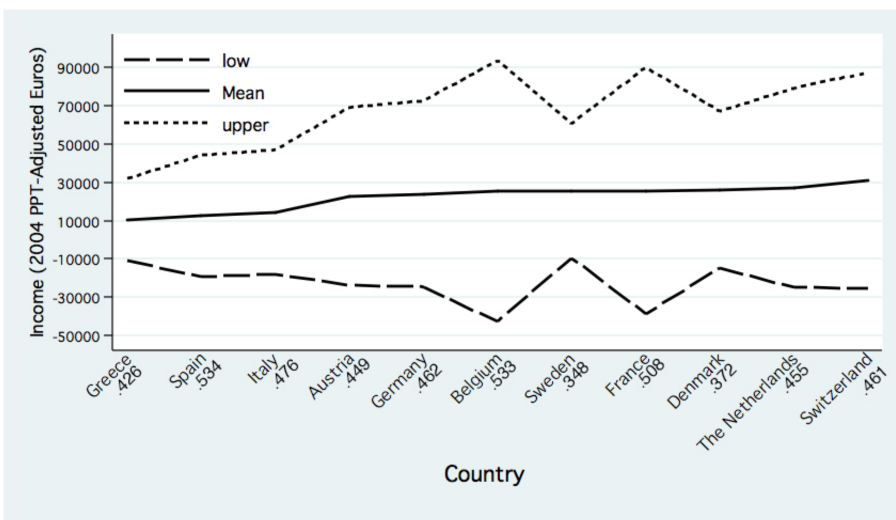

Figure 2: SHARE Mean Gross Equivalised Income, PPP-adjusted, by Country; 95% Confidence Intervals and Gini Coefficients.

SAH scales S_1 to S_4 can be seen for Sweden, by way of example, in Figure 3.

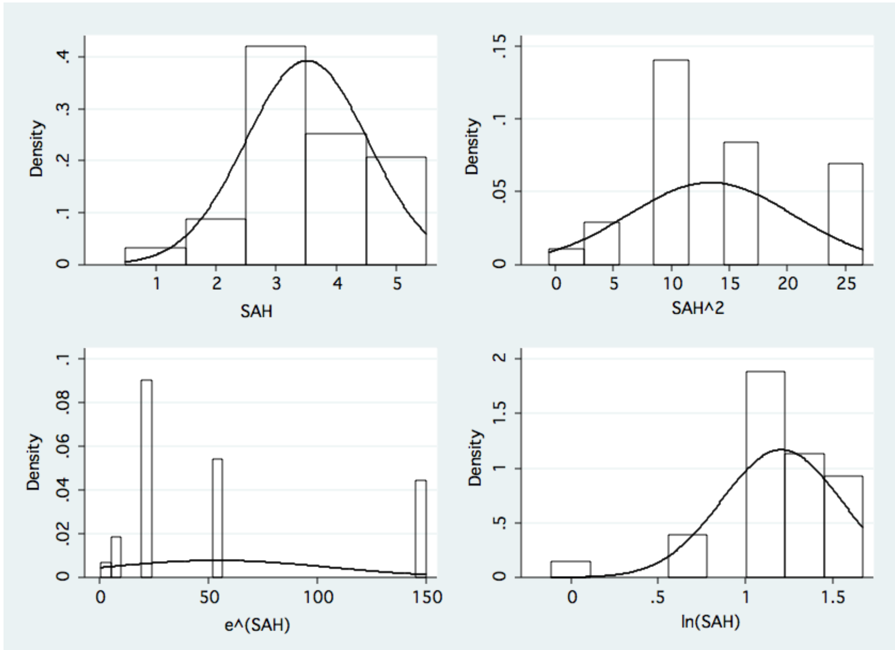

Figure 3: Histograms and Kernel Densities for non-linear cardinal SAH scales, Sweden only.

Note the different scales along both axes in the four graphs. The non-linearity of the transformations is moving the mean/skewness as they occur. One can see that, depending upon the starting point of one country's mean under S_1, subsequent means will move relative to those of other countries under the same transformation also, affecting internationally-comparative measures of inequality.

6 Results

6.1 Decomposed rank-covariance indices

Table 1 contains results from decomposition of the rank-covariance indices in Equation (4). Statistical significance indicated is taken from ordinary regression of the factors of decomposition on continued SAH as described in Section 2.2.

	Austria	Germany	Sweden	The Netherlands	Spain	Italy	France	Denmark	Greece	Switzerland	Belgium
Female	0.0010	0.0002	0.0013**	0.0001	-0.0002**	0.0004**	0.0000	0.0002	0.0017**	0.0013	0.0003**
	39.6924	*1.8971*	*9.0222*	*1.3596*	*-8.1373*	*5.6833*	*-0.3421*	*1.1959*	*25.9217*	*14.1192*	*5.3051*
No Education	0.0004	0.0003**	0.0003*	0.0002**	0.0030**	0.0017**	0.0064**	0.0003	0.0019**	0.0000	0.0005**
	14.9014	*2.7372*	*2.3750*	*3.1819*	*128.1805*	*21.6676*	*60.6987*	*1.3023*	*28.9493*	*0.4477*	*10.7677*
Post-Sec. Educ	0.0032**	0.0066**	0.0075**	0.0064**	0.0024**	0.0045**	0.0111**	0.0089**	0.0094**	0.0095**	0.0053**
	128.8689	*52.3132*	*51.8279*	*127.0028*	*102.6389*	*58.6120*	*105.0850*	*43.5704*	*140.7480*	*101.6204*	*107.1859*
Smoking	0.0008**	0.0003**	0.0011**	0.0003**	-0.0001	0.0001	-0.0001	0.0016**	0.0000	-0.0021**	0.0000**
	32.0694	*2.5687*	*7.9354*	*6.2570*	*-2.3706*	*0.6830*	*-1.1243*	*7.6327*	*0.1032*	*-22.6833*	*0.3511*
Drinking	0.0000	0.0000	0.0011**	0.0007**	0.0014**	0.0014**	0.0003	-0.0004	0.0000	0.0002	0.0003
	-0.6598	*-0.2998*	*7.4323*	*13.1629*	*58.2386*	*18.5600*	*2.4555*	*-1.8032*	*0.0755*	*2.5663*	*5.5668*
Overweight	-0.0001*	0.0008**	0.0001**	0.0002	0.0001**	-0.0002**	0.0001**	-0.0001**	-0.0002**	0.0003	0.0000**
	-3.6118	*6.3188*	*0.5137*	*3.9879*	*5.8447*	*-3.0728*	*0.9074*	*-0.3432*	*-3.6131*	*3.1244*	*0.4319*
Obese	0.0022**	0.0035**	0.0008**	0.0009**	0.0009**	0.0034**	0.0030**	0.0019**	0.0022**	0.0001**	0.0011**
	89.4058	*27.7969*	*5.6755*	*17.5418*	*40.4426*	*43.4709*	*28.6472*	*9.0713*	*32.2491*	*0.9419*	*21.6165*
Retired Young	0.0001	-0.0015**	-0.0044**	0.0002	0.0001	-0.0003	-0.0001	-0.0045	0.0001*	0.0006	0.0001
	4.9598	*-12.1923*	*-30.2790*	*4.6764*	*4.0134*	*-3.7135*	*-1.2740*	*-21.7729*	*1.6786*	*6.0065*	*1.8693*
Age	-0.0119**	0.0064**	0.0157**	-0.0041**	0.0012**	-0.0061**	-0.0041**	0.0308	0.0013**	0.0016**	-0.0023**
	-473.3220	*50.7303*	*108.9940*	*-81.0840*	*51.4424*	*-78.0495*	*-38.4170*	*150.5284*	*20.0255*	*16.7500*	*-46.9052*
Residual	0.0067	-0.0040	-0.0092	0.0002	-0.0065	0.0028	-0.0060	-0.0183	-0.0098	-0.0021	-0.0003
	267.6960	*-31.8701*	*-63.4970*	*3.9137*	*-280.2932*	*36.1590*	*-56.6365*	*-89.3817*	*-146.1378*	*-22.8930*	*-6.1892*
CI	0.0025	0.0126	0.0144	0.0050	0.0023	0.0078	0.0106	0.0205	0.0067	0.0094	0.0050

** Significant at 5%
* Significant at 10%

Table 1: Decomposed concentration indices (*percentage of total CI*) for Empirical CDF Ranks of SAH and Income.

"Retired young" is an indicator of whether or not the individual is less than 65 years of age at the time surveyed. This is to capture the most likely threshhold effect of age, under the assumption that age will affect inequality through sub-groups. Age was included to ease the grid-searching of the optimiser for the copula models, later.[10] However it can interfere with the observed effects of early retirement as a result.

Primary education was collinear with no and post-secondary education, relative to secondary education, and was dropped. Asterisks show statistical significance from the ordinary regressions used to calculate the weights in Equation (6). This is necessary because insignificant factors in the model can still appear to have a higher impact under decomposition, however their relative lack of information makes them poor candidates for any policy response.

The residual Concentration Index was calculated as the remainder between the simple Concentration Index in Equation (1) and the linear combination of the decomposed indices in Table 1.[11] The residual is substantial but not consistently-signed across the SHARE countries. It is driven by common variables: age, in particular, as well as education and obesity, albeit to a lesser extent. Residual indices are also greatest for Austria and Spain, who have the lowest Concentration Index overall, however this relationship is not born out across all of the countries. It is worth noting, though, that all countries have relatively low degrees of association between SAH and income, which affects the degree to which the association that does exist can be explained.

6.2 Copulas

The Frank, Clayton, Gumbel and Gaussian copulas from Equations (9)-(12) were estimated and then compared using Information Criteria to determine the best-fitting specification. The Frank copula performed best according to these criteria; results are contained in Table 2.[12]

	Austria	Germany	Sweden	The Netherlands	Spain	Italy	France	Denmark	Greece	Switzerland	Belgium
Female	0.0889	-0.2676	-0.1742	0.0107	0.0922	0.2695*	0.5212**	0.4471	-0.0836	-0.0660	0.1495*
	0.0787	0.3180	0.3697	0.0874	0.0864	0.1477	0.2491	0.6035	0.1437	0.2244	0.0892
No Education	0.0922	1.4161	14.9777	0.3128	0.1371	0.3123	1.0344**	4.2622	0.1026	-0.5586	-0.1917
	166.4633	3.4792	583.6023	2.0093	0.1004	0.5506	0.3869	143.1876	0.2100	32.6532	0.4267
Post-Sec. Educ	0.1035	0.1700	0.7584	0.0155	0.0590	-0.0024	0.6296*	0.7589	-0.0652	0.1486	0.1253
	0.0818	0.3141	0.4642	0.1622	0.0996	0.3927	0.3362	0.6748	0.1135	0.2441	0.1563
Smoking	-0.0182	0.2550	0.3475	-0.0469	0.1848**	-0.1320	0.5385	-1.4583	-0.0040	0.0432	-0.0076
	0.0899	0.4152	0.4936	0.1241	0.0970	0.1947	0.4039	0.8459	0.1254	0.2607	0.1384
Drinking	0.0847	-0.3439	1.6166	0.0157	0.0714	0.1190	0.1584	-0.7667	-0.2063	0.1627	0.1145
	0.1298	0.6301	2.3891	0.1425	0.0769	0.1454	0.3476	0.9453	0.2413	0.3103	0.1505
Overweight	0.2837**	0.0591	-1.5966**	0.0842	0.0440	-0.2142	-0.1578	-0.3227	-0.0559	0.1954	0.0778
	0.0872	0.3820	0.4267	0.1173	0.0779	0.1633	0.2960	0.8948	0.1151	0.2484	0.1002
Obese	0.3585**	-0.6163	-1.5129**	0.0051	0.0834	0.7764**	-0.1168	-1.4051	0.2470	-0.4716	-0.0555
	0.1003	0.4819	0.5596	0.1192	0.1082	0.2598	0.4137	0.9218	0.2400	0.3089	0.1318
Retired Young	-0.0198	-0.3134	-0.4404	0.0103	0.1694**	0.7811**	-0.8386**	-	0.2908*	-0.4211	0.1585
	0.1032	0.4017	0.5055	0.1797	0.0767	0.2029	0.3128	-	0.1348	0.3299	0.1580
Age	0.0121**	-0.0081	0.0027	0.0137**	0.0153	0.0169**	0.0212**	-0.0088	0.0274**	0.0196	-0.0001
	0.0022	0.0149	0.0188	0.0034	0.0032	0.0050	0.0124	0.0232	0.0045	0.0088	0.0070
Cons	-0.0785	2.5545**	2.7211**	-0.0013	-0.2819**	-0.0002	0.1202	4.6559**	-0.3565**	0.1954	0.7605*
	0.1168	0.9093	1.0223	0.2300	0.1279	0.1997	0.7381	1.6927	0.1334	0.4642	0.4158
θ	0.6695	2.0109	2.2830	0.9377	0.8830	1.2371	1.7369	3.7694	1.0151	1.3680	0.9317

** Significant at 5%
* Significant at 10%

Table 2: Coefficients (*standard errors*) for the Frank copula of Ranks of SAH and Income.

The copula models had routinely unstable hessians, and were estimated via bootstrapping. Even so, and with the inclusion of age as a continuous variable, the indications were that the maximisation was finding local maxima only; the results do not appear to be robust to difference starting values.

Because of this, limited statistical significance is found across the copula models. There is some suggestion however that, at least in some countries, overweight and obesity, as well as potentially related early retirement, has an effect on dependence between continued SAH and income. The effect is not consistently-signed, though - unlike other countries the effect of increasing rates of overweight, obesity and early retirement correlate with decreased dependence. As with the Concentration Indices, dependence was relatively small for all of the countries, making convergence more difficult.

6.3 Quantile Regression

After selecting the Frank copula and jackknifing samples for each country, quantile regression at the median was undertaken. Results for the median and interquartile ranges are contained in Tables 3-5.

	Austria	Germany	Sweden	The Netherlands	Spain	Italy	France	Denmark	Greece	Switzerland	Belgium
Female	0.0095	-0.0634	-0.0081	-0.0973	0.0822	0.0624	0.0424	0.0577	-0.1046	-0.0627	-0.0908
	0.1395	*0.0493*	*0.0515*	*0.0753*	*0.1325*	*0.0761*	*0.0564*	*0.0499*	*0.0696*	*0.1210*	*0.0700*
No Education	3.5231**	0.5009**	1.8815**	1.0030**	0.2590	0.3326*	0.3411**	0.5294**	0.3209**	-0.5241	0.0931
	0.8138	*0.2547*	*0.4029*	*0.3625*	*0.1574*	*0.1891*	*0.0675*	*0.0791*	*0.1196*	*0.8673*	*0.2645*
Post-Sec. Educ	0.2396	0.0649	0.0585	0.2770**	0.7091**	0.6639**	0.4697**	0.0322	0.6877	0.0324	0.1842**
	0.1555	*0.0533*	*0.0562*	*0.0942*	*0.2340*	*0.1370*	*0.0690*	*0.0529*	*0.0903*	*0.1346*	*0.0804*
Smoking	-0.0931	-0.0455	-0.0313	0.1268	-0.0696	0.1751*	0.0745	0.0488	-0.0251	-0.0834	-0.1313
	0.1723	*0.0625*	*0.0700*	*0.0874*	*0.1777*	*0.0940*	*0.0765*	*0.0533*	*0.0783*	*0.1438*	*0.0896*
Drinking	-0.2380	-0.0517	0.0235	0.0002	-0.1462	-0.0918	0.0418	0.0615	0.0875	0.0234	-0.0825
	0.2446	*0.0828*	*0.1842*	*0.0978*	*0.1822*	*0.0880*	*0.0665*	*0.0684*	*0.1284*	*0.1674*	*0.0971*
Overweight	-0.0733	-0.1299**	-0.0763	-0.1034	0.1597	0.0343	-0.0577	-0.0476	-0.1079	-0.1834	-0.0378
	0.1459	*0.0517*	*0.0554*	*0.0799*	*0.1421*	*0.0790*	*0.0579*	*0.0533*	*0.0736*	*0.1295*	*0.0742*
Obese	0.0662	-0.1072	-0.0710	-0.0153	0.0322	-0.0315	0.0058	-0.0488	-0.1365	0.0627	-0.0608
	0.1835	*0.0683*	*0.0791*	*0.1120*	*0.1665*	*0.1012*	*0.0776*	*0.0760*	*0.0943*	*0.1809*	*0.0959*
Retired Young	0.0221	-0.1015*	-0.2597**	-0.2748*	-0.0535	0.0085	-0.0238	0.0849	-0.1523	-0.0379	0.0168
	0.2260	*0.0570*	*0.0901*	*0.1394*	*0.2183*	*0.1219*	*0.0998*	*0.0900*	*0.1223*	*0.2093*	*0.0829*
Age	-0.0128	-0.0096**	0.0157**	0.0093	-0.0079	-0.0010	-0.0025	-0.0023	0.0006	-0.0024	-0.0085**
	0.0114	*0.0010*	*0.0045*	*0.0068*	*0.0104*	*0.0065*	*0.0044*	*0.0040*	*0.0056*	*0.0098*	*0.0014*
Cons	0.8829	0.9897	-0.6150**	-0.3573	0.4208	0.1046	0.1894	0.3822	0.1570	0.4556	0.7238
	0.6793	-	*0.2689*	*0.3999*	*0.6252*	*0.3896*	*0.2554*	*0.2378*	*0.3301*	*0.5757*	-
** Significant at 5%											
* Significant at 10%											

Table 3: Coefficients (*standard errors*) for Quantile regression of jackknifed samples at the lower quartile range

	Austria	Germany	Sweden	The Netherlands	Spain	Italy	France	Denmark	Greece	Switzerland	Belgium
Female	0.0821	-0.0411	0.0127	-0.0091	-0.0404	0.0077	0.0668**	0.0344	-0.0209	-0.0413	0.0328
	0.0676	*0.0263*	*0.0229*	*0.0312*	*0.0483*	*0.0274*	*0.0317*	*0.0274*	*0.0345*	*0.0517*	*0.0358*
No Education	-0.2784	0.4419**	0.9619**	0.5470**	0.0109	0.0495	0.1558**	0.9466**	0.0184	-0.2133	0.0894
	0.6794	*0.1399*	*0.3104*	*0.1499*	*0.0582*	*0.0681*	*0.0379*	*0.0438*	*0.0594*	*0.3335*	*0.1386*
Post-Sec. Educ	-0.0074	0.0198	0.0092*	0.0766*	0.2986**	0.0753	0.1525**	0.0224	0.2174**	-0.0069	0.1470**
	0.0759	*0.0284*	*0.0256*	*0.0393*	*0.0859*	*0.0489*	*0.0390*	*0.0293*	*0.0449*	*0.0581*	*0.0417*
Smoking	-0.0246	0.0159	-0.0048	-0.0027	0.0173	0.0526	-0.0124	0.0011	-0.0529	-0.0241	-0.0118
	0.0845	*0.0336*	*0.0312*	*0.0363*	*0.0642*	*0.0340*	*0.0427*	*0.0292*	*0.0387*	*0.0623*	*0.0462*
Drinking	-0.0786	-0.0561	0.0500	0.0802**	-0.0947	-0.0154	0.0060	-0.0211	0.0777	-0.0721	0.0292
	0.1194	*0.0445*	*0.0818*	*0.0406*	*0.0662*	*0.0317*	*0.0372*	*0.0377*	*0.0635*	*0.0712*	*0.0501*
Overweight	0.0319	-0.0183	-0.0257	-0.0374	0.0321	0.0097	-0.0272	-0.0195	-0.0249	-0.0276	-0.0331
	0.0713	*0.0275*	*0.0247*	*0.0331*	*0.0520*	*0.0283*	*0.0325*	*0.0293*	*0.0363*	*0.0547*	*0.0385*
Obese	0.0486	0.0053	0.0028	0.0108	-0.0088	-0.0387	0.0375	-0.0749*	-0.0537	-0.0070	-0.0051
	0.0898	*0.0366*	*0.0355*	*0.0464*	*0.0609*	*0.0362*	*0.0434*	*0.0419*	*0.0463*	*0.0774*	*0.0497*
Retired Young	0.0171	0.0096	-0.0917**	-0.0103	0.0111	-0.0325	0.0158	0.0520	-0.0020	-0.0447	0.0886**
	0.1083	*0.0305*	*0.0401*	*0.0572*	*0.0804*	*0.0439*	*0.0555*	*0.0510*	*0.0603*	*0.0901*	*0.0430*
Age	-0.0035	-0.0003	0.0063**	-0.0007	-0.0010	0.0031	-0.0018	-0.0007	0.0023	0.0034	-0.0038**
	0.0055	*0.0005*	*0.0020*	*0.0028*	*0.0038*	*0.0024*	*0.0025*	*0.0023*	*0.0027*	*0.0042*	*0.0007*
Cons	0.2339	0.2610	-0.1185	0.1492	0.1279	-0.0381	0.2135	0.3337	0.0026	0.0326	0.2769
	0.3306	-	*0.1168*	*0.1629*	*0.2254*	*0.1413*	*0.1456*	*0.1329*	*0.1579*	*0.2429*	-

** Significant at 5%
* Significant at 10%

Table 4: Coefficients (*standard errors*) for Quantile regression of jackknifed samples at the median

	Austria	Germany	Sweden	The Netherlands	Spain	Italy	France	Denmark	Greece	Switzerland	Belgium
Female	0.2192*	-0.1208*	0.0545	-0.0758	-0.1020	0.0306	0.1469**	0.1178	0.1127	-0.0468	0.1281*
	0.1310	*0.0711*	*0.0641*	*0.0892*	*0.1362*	*0.0931*	*0.0515*	*0.0931*	*0.0756*	*0.1360*	*0.0759*
No Education	0.5199	0.2145	1.5314**	0.8008*	-0.0715	-0.1207	-0.0266	1.2922**	-0.0544	0.5301	0.1867
	0.7681	*0.3654*	*0.5074*	*0.4299*	*0.1649*	*0.2275*	*0.0614*	*0.1490*	*0.1293*	*0.9804*	*0.2876*
Post-Sec. Educ	0.1145	-0.0759	-0.0577	-0.0288	-0.0766	0.0245	0.0993	-0.0190	0.1139	-0.0312	0.3412**
	0.1480	*0.0756*	*0.0716*	*0.1142*	*0.2415*	*0.1640*	*0.0631*	*0.1006*	*0.0987*	*0.1551*	*0.0883*
Smoking	-0.1312	-0.0001	0.0572	-0.0971	0.1108	0.1204	-0.1127*	0.0239	-0.0460	0.0979	0.1098
	0.1656	*0.0902*	*0.0874*	*0.1043*	*0.1801*	*0.1137*	*0.0682*	*0.0994*	*0.0849*	*0.1655*	*0.0978*
Drinking	0.0110	-0.1892	0.0466	0.1280	-0.2280	-0.0504	0.0084	-0.0023	0.1957	-0.0999	0.0859
	0.2296	*0.1200*	*0.2251*	*0.1158*	*0.1853*	*0.1067*	*0.0600*	*0.1282*	*0.1378*	*0.1885*	*0.1063*
Overweight	0.3751**	-0.0478	0.0147	-0.0309	0.0301	0.0436	0.0100	0.0063	0.0331	0.0718	-0.0100
	0.1380	*0.0744*	*0.0698*	*0.0947*	*0.1470*	*0.0952*	*0.0525*	*0.0999*	*0.0792*	*0.1453*	*0.0815*
Obese	0.4417**	-0.1134	-0.0132	-0.1507	0.1021	-0.0357	-0.0554	-0.0715	0.0841	-0.1898	0.0742
	0.1742	*0.0983*	*0.0999*	*0.1331*	*0.1721*	*0.1217*	*0.0700*	*0.1414*	*0.1008*	*0.2035*	*0.1045*
Retired Young	0.0026	0.1783**	-0.1343	-0.0908	0.2814	0.0138	-0.0729	-0.0288	-0.0670	0.0775	0.3056**
	0.2098	*0.0804*	*0.1122*	*0.1603*	*0.2219*	*0.1447*	*0.0881*	*0.1739*	*0.1272*	*0.2389*	*0.0915*
Age	0.0069	-0.0011	0.0136**	-0.0008	0.0087	0.0066	-0.0037	-0.0027	0.0167	0.0042	-0.0053**
	0.0112	*0.0014*	*0.0056*	*0.0078*	*0.0106*	*0.0080*	*0.0041*	*0.0076*	*0.0058*	*0.0110*	*0.0015*
Cons	-0.7501	0.3588	-0.5985**	0.2753	-0.5914	-0.3107	0.3933	0.4409	-0.9962**	-0.1022	0.1245
	0.6724	-	*0.3321*	*0.4574*	*0.6355*	*0.4819*	*0.2391*	*0.4460*	*0.3415*	*0.6354*	-
** Significant at 5%											
* Significant at 10%											

Table 5: Coefficients (*standard errors*) for Quantile regression of jackknifed samples at the upper quartile range

Like the copula results in Table 2, there is substantially less statistical significance in the quantile regression results, relative to the decomposed Concentration Index. Quantile regression has the advantage, though, of demonstrating different effects according to quantiles of combinations of the regressors: the marginal effects at higher and lower degrees of dependence overall.

The results show some consistency in terms of the effect of education; countries differ, however, in whether no education or post-secondary education is pre-eminent, and only France and, to a lesser extent the Netherlands, show significant effects for both. Lifestyle effects are not consistent with the copula results. At the 75% quartile, gender, overweight and obesity become significant in some countries.

7 Discussion

The residual Concentration Index in Table 1 illustrates one of the problems when attempting to explain income-related inequalities in health by factors. Under decomposition, one is concerned with the distribution of income across the factors of health, weighted according to the degree to which they explain health. In terms of statistical significance, however, only with regards to SAH can it be known. Moreover, the residual indices were substantial in several countries, leaving the issue of factorisation somewhat unresolved.

This is not a newly-discovered problem. Amongst other things, one of the shortcomings with attempting to isolate the factors of income-related inequalities in health is the specification: although this model satisfied specification for all of the countries, omitted variables is unavoidable; in particular, income (or income-related socioeconomic status) cannot be included because (i) the causality is complex, and difficult to accommodate in cross-sectional data (Jones and L*ó*opez Nicol*á*s 2004), and (ii) including income directly into the equation for SAH risks introducing inequalities in the distribution of income into the measured distribution of health, thereby contaminating what is actually being measured.

Table 2 illustrates the direct marginal effect of each covariate on the dependence parameter. In doing so, it provides new information on how best to respond to socioeconomic inequalities in health: using direct marginal effects, one can gauge those factors that are (i) statistically significant, (ii) economically meaningful and (iii) subject to policy. With respect to the Concentration Index, the issue of inclusion socioeconomic indicators directly remains open: as income is one of the margins of the bivariate model, including it as a covariate in dependence is not feasible.

The balance of the evidence across the three approaches suggests that socioeconomic inequalities in SAH are explained by education, on the one hand, and overweight/obesity on the other. Since these are linked with both health and income it is not necessarily a surprise to see this narrative form.

Future analysis using multivariate, rather than bivariate, models, can consider this evidence: education could itself be included with health and income in a multivariate distribution function. This parsing of dependence would follow the jackknife/quantile regression approach, sorting dependence into levels of education. So, too, would consideration of more complex copulas: the quantile regression results suggest that multi-parameter, asymmetric and/or tail-dependent copulas could provide more information than single-parameter, symmetric copulas.

8 Conclusion

In this paper, I presented results for concentration indices as well as copulas and quantile regression of jackknifed samples for an international comparison of income-related inequalities in SAH using data on retirees in 11 European countries. The results show that parameterisation of dependence in the copula generates different information on size and statistical significance of the factors of inequality when measured directly, rather than indirectly through Generalised Linear Models for health.

Although the performance and results of the methods used varied, the evidence suggests the socioeconomic inequalities vary the most at different levels of education, in accordance with the human capital theory of investment in education and health. Quantile regression with jackknifed pseudo-data further suggested that multi-parameter and asymmetric copula functions should be used to gain further information on socioeconomic inequalities in health.

Acknowledgements

Help from Murray Smith, Jason Hockenberry and Silvia Balia was invaluable in constructing this paper, although any errors are mine alone. So too was the advice of Phillip Clarke and David van der Goes. This work has been presented at the Centre for Health Economics Research Evaluation at the University of Technology, Sydney; at the Melbourne Institute; at the Centre for Aboriginal Economic Policy Research at the Australian National University; and at the 2009 Scottish Economics Society Conference. All of the participants involved are thanked for their input.

Notes

[1]Concordance is preserved because the original ordinal scale can always be recreated from a continued discrete random variable.

[2]STATA, for example, will carry this out automatically upon omission of the *equal* option when using the *cumul* command.

[3]Taken from Podder, N. 1993. The disaggregation of the Gini coefficient by factor components and its application to Australia, Review of Income and Wealth, 39(1): pp 52-53.

[4]Note that $v = G(Y) = Rank(Y) = R$, as in previous models.

[5]This will keep the copula and quantile regression estimates comparable, however it should be noted that, as Archimedean-class copulas, the parameters of the Frank, Clayton, Gumbel and Gaussian copulas can all be expressed as either Kendall's τ or Spearman's ρ (Nelsen 2006). Not regarding this comparison, the copula need not be used at all: quantile regression of jackknifed $\hat{\tau}_i$ in $i \in (1, .., n)$ replications is equivalent.

[6]http://www.share-project.org/. This paper uses data from SHARE Waves 1 & 2, as of December 2008. SHARE data collection in 2004-2007 was primarily funded by the European Commission through its 5th and 6th framework programmes (project numbers QLK6-CT-2001- 00360; RII-CT- 2006-062193; CIT5-CT-2005-028857). Additional funding by the US National Institute on Aging (grant numbers U01 AG09740-13S2; P01 AG005842; P01 AG08291; P30 AG12815; Y1-AG-4553-01; OGHA 04-064; R21 AG025169) as well as by various national sources is gratefully acknowledged (see http://www.share-project.org for a full list of funding institutions).

[7]This is due to missing information on health and income in the first wave.

[8]http://www.mea.uni-mannheim.de

[9]Although there is reason to expect endogeneity amongst these variables, it is not considered during the ensuing analysis, mostly to avoid encumbering unnecessarily the analysis and discussion of the methods. The effect is likely to be common across all countries for this data (e.g. age and education, health and income, for example).

[10]Age is a continuous variable, and can smooth over relatively empty cells made up of dummy variables in a model such as this. Intersections of retirement, overweight/obesity and education levels, in particular, were consistently under-populated.

[11]The intercept from the model explaining continued SAH is remaindered to the residual CI as well.

[12]results from the other copulas can be obtained from the author.

References

Allison, R. A., Foster, J. E. 2004. Measuring health inequality using qualitative data. Journal of Health Economics. 23: 505-524.

Bogaerts, K., Lesaffre, E. 2008. Modeling the association of bivariate interval-censored data using the copula approach. Statistics in Medicine. 27(29): 6379-6392.

Bommier, A., Stecklov, G. 2002. Defining health inequality: why Rawls succeeds where social welfare theory fails. Journal of Health Economics. 21: 497-513.

Börsch-Supan, A., Brugiavini, A., Jürges, H., Kapteyn, A., Mackenbach, J., Siegrist, J., Weber, G. (eds.) 2005. Health, ageing and retirement in Europe: first results from the Survey of Health, Ageing and Retirement in Europe. Mannheim Research Institute for the Economics of Aging, University of Mannheim.

Clarke, P.M., Gerdtham, U.-G., Johannesson, M., Bingefors, K., Smith, L., 2002. On the measurement of relative and absolute income-related health inequality. Social Science & Medicine. 55(11): 19231928.

Clayton, D.G. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika. 65(1): 141-151.

Denuit, M., Lambert, P. 2005. Constraints on concordance measures in bivariate discrete data. Journal of Multivariate Analysis. 93(1): 40-57.

Erreygers, G. 2006. Correcting the Concentration Index. University of Antwerp. Antwerp, Belgium.

Erreygers, G. 2009a. Correcting the Concentration Index. Journal of Health Economics. 28: 504-515.

Erreygers, G. 2009b. Correcting the Concentration Index: a reply to Wagstaff. Journal of Health Economics. 28: 521-524.

Frank, M.J. 1979. On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. Aequationes Mathematica, 19: 194-226.

Good, P. 2002. Extensions of the concept of exchangeability and their applications. Journal of Modern Applied Statistical Methods. 1: 243-247.

Grossman, M. 1973. The Correlation Between Health and Schooling. NBER Working Papers 0022. National Bureau of Economic Research, Inc.

Hicks, J., Allen, R.G.D. 1934. A reconsideration of the theory of value I. Economica. 1(1): 52-75.

Hollander, H. 2001. On the validity of utility statements: Standard theory versus Duesenberrys. Journal of Economic Behavior and Organization. 45(3): 227-249.

International Society for Equity in Health. 2001. Working definitions http://www.iseqh.org

Joe, H. 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analysis. 942: 401-419.

Jones, A.M., L*ó*opez Nicol*á*s, A. 2004. Measurement and explanation of socioeconomic inequality in health with longitudinal data. Health Economics 13(10): 1015-1030.

Kakwani, N., 1980. *Income Inequality and Poverty.* World Bank, New York.

Koenker, R. 2005. Quantile Regression. Cambridge University Press, New York.

Mesfioui, M., Tajar, A. 2005. On the properties of some nonparametric concordance measures in the discrete case. Journal of. Nonparametric Statistics. 17(5): 541-554.

Nelsen, R.B. 2006. An introduction to copulas, 2nd Ed. Springer Verlag, New York.

Nikoloulopoulos, A.K., Karlis, D. 2008. Multivariate logit copula model with an application to dental data. Statistics in Medicine, 27(30): 6393-6406.

O'Donnell, O., van Doorslaer, E., Wagstaff, A., Lidelow, M. 2008. Analysing Health Equity Using Household Survey Data: A Guide To Techniques And Their Implementation. Health Financing Policy, World Health Organization, Geneva.

Podder, N. 1993. The disaggregation of the Gini coefficient by factor components and its application to Australia. Review of Income and Wealth, 39(1): 51-61.

Quinn, C. 2009. Scaling concentration indices with ranked Self-Assessed Health. Mimeo.

Shao, J., Tu, D. 1995. The Jackknife and Bootstrap. Springer Verlag, New York.

Smith, M.D. 2007. Invariance theorems for Fisher information. Communications in Statistics, Theory and Methods. 36(12): 2213-2222.

van Doorslaer, E., Jones A. M. 2003. Inequalities in self-reported health: validation of a new approach to measurement Journal of Health Economics. 221: 61-87.

Wagstaff, A. 2005. The bounds of the Concentration Index when the variable of interest is binary, with an application to immunization inequality. Health Economics 14(4): 429432.

Wagstaff, A. 2009. Correcting the Concentration Index: a comment. Journal of Health Economics. 28: 516-520.

Wagstaff, A., van Doorslaer, E., Watanabe, N. 2003. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. Journal of Econometrics. 112: 207-223.

Wagstaff, A., Paci, P., van Doorslaer, E. 1991. On the measurement of inequalities in health. Social Science and Medicine. 335: 545-557.

Yan, Y. 2007. Enjoy the joy of copulas: with a package copula. Journal of Statistical Software. 21(4): 1-21.