DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 3/2010

On the life course perspective in income related health inequalities: A semiparametric approach

by

Martin Siegel

Karl Mosler

October 2010

DISKUSSIONSBEITRÄGE ZUR STATISTIK UND ÖKONOMETRIE

SEMINAR FÜR WIRTSCHAFTS- UND SOZIALSTATISTIK
UNIVERSITÄT ZU KÖLN

Albertus-Magnus-Platz, D-50923 Köln, Deutschland

DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 3/2010

On the life course perspective in income related health inequalities: A semiparametric approach

by

Martin Siegel[1,2]

Karl Mosler[3]

October 2010

Abstract

Compared with the broad supply of literature measuring socioeconomic gradients in the distribution of health, only little is known about the life course perspective regarding income related inequalities. This article combines the renowned concentration index approach with semiparametric estimation techniques to derive a new varying index of inequality that copes without a priori sample stratification. We illustrate the power of this new index using health data drawn from the German microcensus and find support for the age as leveler hypothesis. Our index suggests that significant inequalities to the detriment of the deprived increase over the working life and reach their maximum around the age for retirement. We find no significant inequalities for adolescents and elder people.

[1]Corresponding author, email: martin.siegel@uk-koeln.de
[2]Institute of Health Economics, University Hospital of Cologne, Cologne, Germany
[3]Department of Economic and Social Statistics, Albertus-Magnus-University of Cologne, Cologne, Germany

1 Introduction

The existence of socioeconomic gradients in the distribution of health to the detriment of the deprived is firmly established among health economists (Balia and Jones, 2008; Erreygers, 2009; Humphries and van Doorslaer, 2000; Kakwani et al., 1997; van Doorslaer et al., 1997; van Doorslaer and Koolman, 2004; van Doorslaer et al., 2004; Wagstaff et al., 1991; Wagstaff and van Doorslaer, 2000; Wagstaff et al., 2003; Wagstaff, 2005). Little is known, however, about the mechanisms through which different socioeconomic factors affect health status and its distribution over the life course (van Kippersluis et al., 2010, 2009). Adding the life course perspective supports, for instance, the notion that labor force participation contributes substantially to the socioeconomic gradient in health in the U.S. (Case and Deaton, 2005), Great Britain (Banks et al., 2007) and the Netherlands (van Kippersluis et al., 2010).

In this article, we employ the two common hypotheses explaining why disparities in health may differ over the life course: the accumulation hypothesis and the age as leveler hypothesis. According to the former, low socioeconomic status and lacking resources are associated with less healthy lifestyles. Adverse psychosocial effects and health behaviors accumulate over the adult life course, eventually resulting in highest income related health inequality in the oldest age (Kim and Durden, 2007; Lynch, 2003; Prus, 2004; Ross and Wu, 1996; Willson et al., 2007). Conversely, the age as leveler hypothesis suggests that adverse effects and psychosocial stress exist over the working life but relief with retirement. One would hence expect the greatest income related health disparities around the common age for retirement (Case and Deaton, 2005; Deaton and Paxson, 1998; Elo and Preston, 1996; Herd, 2006; Kim and Miech, 2009; Kunst and Mackenbach, 1994).

A broad supply of literature compares the (predicted) health status over the lifetime for distinct socioeconomic groups. In a recent paper, van Kippersluis et al. (2010) compare the development of self assessed health over the life course using cross sectional data; and Mirowsky and Ross (2005) investigate differences in the predicted health status between educational levels. This method is well suited to describe the development of the groups' health, however, defining socially advantaged and disadvantaged groups requires an a priori judgement. An alternative to measure health inequalities may be the concentration index (Kakwani et al., 1997; Wagstaff et al., 1991). Although it requires some ranking, say, by income, its computation does not require predefined socioeconomic groups. Wagstaff et al. (2003) propose the decomposition of the concentration index to measure the contribution of demographic and socioeconomic characteristics to total health inequality. Using the marginal effects of these variables on the health outcome, they compute the respective elasticities and rewrite the health concentration index as the sum of the concentration indices of the explaining variables weighted by their respective elasticities. From the decomposition approach, one may infer how health inequality would change if, say, no demographic effects were present. It does, however, not allow comparisons of inequalities across age groups. Using data from eleven European countries, van Kippersluis et al. (2009) define age cohorts and compute batteries of concentration indices for each country. Since the bounds of the concentration index depend on minimum, maximum and mean of the respective variable, they correct their indices following Erreygers (2009) to assure comparability between age groups and countries. Their graphical comparisons support the accumulation hypothesis for most countries. Conversely, their results favor the age as leveler hypothesis for France, Germany and the U.K.

In this paper, we introduce a varying inequality index for dichotomous health variables that copes without a priori sample stratification. Using the semiparametric varying coefficient model (Hastie and Tibshirani, 1993; Li et al., 2002) and applying a Nadaraya-Watson estimator with local bandwidth selection, we propose a semiparametric extension of the convenient regression approach (Kakwani et al., 1997; Lerman and Yitzhaki, 1989). Our measure of health is the question 'Did you experience illness during the last four weeks including chronic illness?'. We adjust our varying index using a correction formula for binary variables (Wagstaff, 2005) with local estimates of the mean. As the residual term is likely to be heteroscedastic and serially correlated (Wildman, 2003), we compute confidence bands using locally estimated heteroscedasticity and autocorrelation consistent standard errors (Newey and West, 1987). To take the sample variability of the mean into account, we approximate the local standard error using the δ method (Rao, 1965).

Our aim is to refine the approach by van Kippersluis et al. (2009). Using semiparametric kernel smoothing and a locally chosen bandwidth allows us to estimate the functional relationship between the concentration index and age consistently.

It has been argued that cross sectional data is not suitable for causal analyses of life course perspectives concerning health inequalities (Prus, 2004), however, alike van Kippersluis et al. (2009), our objective is to describe the differences in income related health inequalities across age cohorts.

The remainder of this paper is organized as follows: in the next section, we present our data and variables. We review the concentration index in section 3 and present the varying coefficient model in section 4. In section 5, we introduce our semiparametric concentration index approach and give computational details in section 6. We illustrate the empirical results in section 7 and give a brief discussion in section 8.

2 Data and variables

Data for the empirical application were drawn from the German microcensus (Mikrozensus) conducted by the Federal Statistical Office (Statistisches Bundesamt). The microcensus is Europe's largest annual country-wide survey with one percent of the German households (approximately 820,000 individuals) being interviewed. Households are included in four consecutive surveys and 25 percent of the households are replaced each year. In 2005, the vast majority of interviews was conducted by trained staff as face to face interviews. Answers were recorded directly into the data collection software. The rate of self-fillers was approximately twelve percent. The microcensus comprises an annually surveyed socioeconomic module for which response is mandatory. A health related part for which responding was voluntary was included in 2005. Due to sample size and mandatory response, the German microcensus can be seen as one of the most representative samples available (see FSO, 2006 or Reeske et al., 2009).

We use the the scientific use file (SUF) available for non-profit research organizations for an empirical illustration. The SUF comprises a randomly drawn subsample of approximately 70% ($N = 477,239$) of the German microcensus (for a technical report, see Lechert and Schimpl-Neimanns, 2007). Inverse probability weights accounting for regional, age and sex specific composition of the sample are provided by the Federal Statistical Office, see Lechert and Schimpl-Neimanns (2007).

We measure health outcome via the question "have you been ill (including chronic diseases) or injured by an accident during the last four weeks" with possible answers "yes, ill", "yes, injured", "no" or "no statement". We restrict the analysis to having been ill and generate a binary variable with outcome 1 for "yes, sick". The options "no" and "yes, injured" were both treated as "not ill" and hence coded as 0. Considering zero household income as non response, we use the modified OECD equivalence scale to compute net equivalent household income. Equivalence weights are assigned as follows: 1 for the first adult, 0.5 for each additional person aged 14 or older and 0.3 for children younger than 14 (see e.g. van Doorslaer et al. 2004; van Kippersluis et al. 2009). We removed $92,458$ individuals from the sample owing to missing information (no statement or zero income); leaving us $N = 384,781$ observations for the empirical analysis.

3 Measuring inequality

Using the covariance approach in Lerman and Yitzhaki (1989), Kakwani et al. (1997) present the regression formula for the concentration index $CI = \frac{2}{N\mu_y}\sum_{i=1}^{N} y_i r_i - 1$ such that $CI = \hat{\tilde{\beta}}_1$ for the regression of a transformed health variable $\tilde{y}$ on the weighted fractional rank r:

$$\frac{2\,\sigma_r^2}{\mu_y} y_i = \tilde{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 r_i + \epsilon_i \tag{1}$$

The index $i = 1, ..., N$ denotes the observation for the i-th individual, y is the (untransformed) health variable with mean μ_y and σ_r^2 is the variance of r. To obtain representative results, weighted regression techniques have to be applied (van Doorslaer and Koolman, 2004). Including sample weights w_i, the weighted fractional rank of an individual i ranked in ascending order by a socioeconomic status variable s is calculated as

$$r_i = \sum_{j=1}^{i} w_j - \frac{w_i}{2} \tag{2}$$

with $\sum_i w_i = 1$ (see Lerman and Yitzhaki 1989; Kakwani et al. 1997). Obviously, the concentration index computed for the socioeconomic status variable s used for ranking, say $y = s =$ income, would be the well known Gini index.

For binary variables, the maximum extent of inequality and therefore the bounds of the concentration index depend on the the variable's mean, $|CI_{max}| \leq 1 - \mu_y$ (see Wagstaff, 2005; Erreygers, 2009). For an intuitive explanation, first assume a constant equal to 1. With no difference between people concentration among rich or poor is evidently impossible; the concentration index equals zero. Now consider, say, 20 percent ones and 80 percent zeros. Ordering the variable by itself would give the Gini index of 0.8; the largest possible concentration. Wagstaff (2005) proposes a corrected concentration index (Wagstaff index)

$$W = \frac{CI}{1 - \mu_y}\,. \tag{3}$$

To obtain the standard error of the concentration index, Kakwani et al. (1997) and Wildman (2003) argue that it is not sufficient to estimate the standard error of $\tilde{\beta}_1$ from equation (1). The error terms are likely to exhibit autocorrelation and, due to the sample variability of $\hat{\mu}_y$, the standard error $\sigma_{\tilde{\beta}}$ of $\tilde{\beta}$ may not be seen as a proper estimate for σ_{CI}. They propose to estimate β_0 and β_1 from the untransformed variable y,

$$y_i = \beta_0 + \beta_1 r_i + \zeta_i\,. \tag{4}$$

Considering the concentration index as a nonlinear combination of β_0 and β_1,

$$CI = \frac{2\,\sigma_r^2}{\mu_y}\,\beta_1 \equiv \frac{2\,\sigma_r^2}{\beta_0 + \mu_r\beta_1}\,\beta_1, \tag{5}$$

with $\beta_0 + \beta_1\mu_r$ in place of μ_y, the variance can be approximated using Rao's δ method (see Rao, 1965). This yields

$$\sigma_{CI}^2 \approxeq 4\sigma_r^4\,\frac{\beta_0^2\sigma_{11} + \beta_1^2\sigma_{22} - 2\,\beta_1\beta_0\sigma_{12}}{(\beta_0 + \beta_1\mu_r)^4}\,, \tag{6}$$

where the σ_{ij} are the ijth elements from the covariance matrix Ω of $\hat{\beta}$. Kakwani et al. (1997) pointed out that the covariance matrix from the OLS regression is not wholly accurate because the error terms ζ_i are not independent from each other and provide an estimator which takes the serial correlation into account. However, this approach has not been extended to weighted regressions (as necessary for weighted samples). We follow Wildman (2003) who proposes using the order of r_i in place of time to compute a heteroscedasticity and autocorrelation consistent Newey-West covariance matrix (see Newey and West, 1987; White, 1980 for computational details). Using the approach from equation (5) to approximate the standard error of the Wagstaff index W, one may rewrite (3) as

$$W = \frac{CI}{1-\mu_y} \equiv \frac{2\,\sigma_r^2}{(\beta_0+\mu_r\beta_1)\,(1-\beta_0-\mu_r\beta_1)}\beta_1. \tag{7}$$

Applying the δ method to equation (7) and performing some straightforward algebra yields

$$\begin{aligned} Var(W) \;\approx\; & \frac{1}{36\left(\beta_0+\frac{1}{2}\beta_1\right)^4\left(1-\beta_0-\frac{1}{2}\beta_1\right)^4} \times \left[\beta_0^2\sigma_{22}\left(\left(1-\beta_0\right)^2-\frac{1}{2}\beta_0\beta_1^2\right)\right. \\ & +\beta_1^2\sigma_{11}\left(1-4(1-\beta_1)\beta_0-2\beta_1+4\beta_0^2+\beta_1^2\right) \\ & +\sigma_{12}\beta_0\beta_1\left(-2+6\beta_0-4\beta_0^2+\beta_1^2+2\beta_1-2\beta_0\beta_1\right) \\ & \left.+\frac{1}{16}\beta_1^3\left(8\beta_1\sigma_{12}+\beta_1\sigma_{22}-8\sigma_{12}\right)+\frac{1}{2}\sigma_{22}\beta_0\beta_1^2\right]. \end{aligned}$$

Note that we replaced mean and variance of the weighted fractional rank with their theoretical values $\mu_r = 0.5$ and $\sigma_r^2 = 1/12$, respectively.

4 Varying coefficient models

In the framework of varying coefficient models, Li et al. (2002) proposed a semi-parametric smooth coefficient model based on locally weighted least squares regression. With X denoting the regressor matrix and y the dependent variable, the elements of the coefficient for vector $(\beta_0, ..., \beta_Q)$ are modeled as smooth functions of another regressor z:

$$y_i = \beta_0(z_i) + \sum_{q=1}^{Q} \beta_q(z_i)x_{q,i} + \epsilon_i \tag{8}$$

This model can be estimated using nonparametric smoothing techniques (see Li et al. (2002) and Hastie and Tibshirani (1993)) as

$$\beta(z) = (E(X'X \mid z))^{-1} \times E(X'y \mid z). \tag{9}$$

As an estimator for $\beta(z)$, Li et al. (2002) have introduced

$$\hat{\beta}(z) = \left[\sum_{i=1}^{N} k_{h_z}(u_i)\, X_i' X_i \right]^{-1} \times \left[\sum_{i=1}^{N} k_{h_z}(u_i)\, X_i' y_i \right], \tag{10}$$

where $u_i = z_i - z$, $k_{h_z}(u_i) = \frac{K_{h_z}(u_i)}{\sum_{j=1}^{N} K_{h_z}(u_j)}$, $K(\cdot)$ is a proper kernel function, and $X_i = \begin{pmatrix} 1 & x_{1,i} & \dots & x_{Q,i} \end{pmatrix}$ denotes the i-th row of X. They have shown that $\hat{\beta}(z)$ asymptotically follows a normal distribution, i.e. $\sqrt{N\, h_z}\left(\hat{\beta}(z) - \beta(z)\right) \backsim N(0, \Omega(z))$ where h_z denotes the bandwidth parameter. The covariance matrix $\Omega(z)$ can therefore be written as

$$\Omega(z) = \left[f(z)\ E(X'X \mid z)\right]^{-1} \Phi_0(z)\ \left[f(z)\ E(X'X \mid z)\right]^{-1} \tag{11}$$

with $\Phi_0(z) = [f(z)\ E(X'X\,\sigma_\epsilon^2(z) \mid X,\, z)\ \ \|K_2^2\|]$ and $\sigma_\epsilon^2(z) = E(\epsilon_i^2 \mid X, z)$.

5 A semiparametric inequality index

Combining the weighted regression approach for the concentration index (1) with the varying coefficient model (8), the proposal for a semiparametric concentration index can be written as

$$2\,\frac{\sigma_r^2(z_i)}{\mu_y(z_i)}\, y_i = \tilde{y}_i = \tilde{\beta}_0(z_i) + \tilde{\beta}_1(z_i)\, r_i(z_i) + \epsilon_i \tag{12}$$

with $CI(z) = \tilde{\beta}_1(z)$. Note that if y is the social status variable, equation (12) works as a semiparametric Gini-index. The local mean $\mu_y(z)$ of y given z can be estimated nonparametrically. The weighted fractional rank $r_i(z_i)$ has to be written as a function of z_i for two reasons. Intuitively, when estimating a varying concentration index, one will be interested in the observable inequality given z. Hence, taking into account all subjects in the sample for computing r_i regardless of their individual values of z seems misleading. Technically, the condition from equation (2) that the sum of the regression weights has to equal 1 can only be fulfilled if the weighted fractional rank is computed using only those individuals included in the local regression and incorporating the kernel smoothing weights, i.e.

$$r_i(z) = \left(\sum_{j=1}^{i} w_j\left(z\right)\, k_{h_z}\left(u_j\right) - \frac{w_i\left(z\right)\, k_{h_z}\left(u_i\right)}{2} \right) \mathcal{I}\left(\left|u_i\right| \leq h_z\right)\,, \tag{13}$$

where $\mathcal{I}\left(\cdot\right)$ denotes an indicator function and the $w_i(z)$ have been rescaled such that $\sum_{i=1}^{N} w_i\left(z\right)\, k_{h_z}\left(u_i\right) = 1$. Consequently, the variance σ_r^2 of the weighted fractional rank has to be written as a function of z. If observations are sparsely distributed around a given z, the number of individuals used may become rather small and as a consequence $\sigma_r^2(z)$ may differ from its theoretical asymptotic value $\frac{1}{12}$. We therefore compute $\hat{\sigma}_r(z)$ from the data.

As the maximum possible extent of inequality depends on the mean, we correct the semiparametric concentration index in order to obtain comparable results throughout the support of z (see Wagstaff, 2005; Erreygers, 2009). Analogously to the Wagstaff index, the proposal for a semiparametric inequality index is a pointwise correction of the semiparametric concentration index from (12) as in equation (3), using the local mean of y:

$$I(z) = \frac{CI(z)}{1 - \mu_Y(z)}.$$

To obtain confidence bands for the semiparametric concentration index, its standard error needs to be estimated. One may follow the approach by Kakwani et al. (1997) and first estimate equation (4) and the respective covariance matrix semiparametrically. Then, confidence bands can be computed using the nonlinear transformation from (5). Using local values for $\hat{\beta}(z)$ and $\hat{\Omega}_{hac}(z)$, equation (6) can be rewritten as

$$\sigma_{CI}^2(z) \;\; \approxeq \;\; 4\sigma_r^4(z)\, \frac{\beta_0^2(z)\,\sigma_{11}(z) + \hat{\beta}_1^{\;2}(z)\,\sigma_{22}(z) - 2\,\beta_1(z)\,\beta_0(z)\,\sigma_{12}(z)}{\left(\beta_0(z) + \beta_1(z)\mu_r(z)\right)^4}\,.$$

for pointwise standard errors with $\sigma_{ij}(z)$ denoting the heteroscedasticity-and-autocorrelation-consistent standard errors from equation (15). Pointwise confidence bands for the semiparametric inequality index can then be easily computed by applying the δ-method to

$$I(z) \equiv \frac{2\,\sigma_r^2(z)}{(\beta_0(z) + \mu_r(z)\beta_1(z))\,(1 - \beta_0(z) - \mu_r(z)\beta_1(z))}\beta_1(z)$$

which yields the above shown result as a function of z.

6 Estimation

Applying the Nadaraya-Watson estimator to obtain consistent estimates, we account for sample weights and use the estimator (10) for $\beta(z)$ with the transformed health variable,

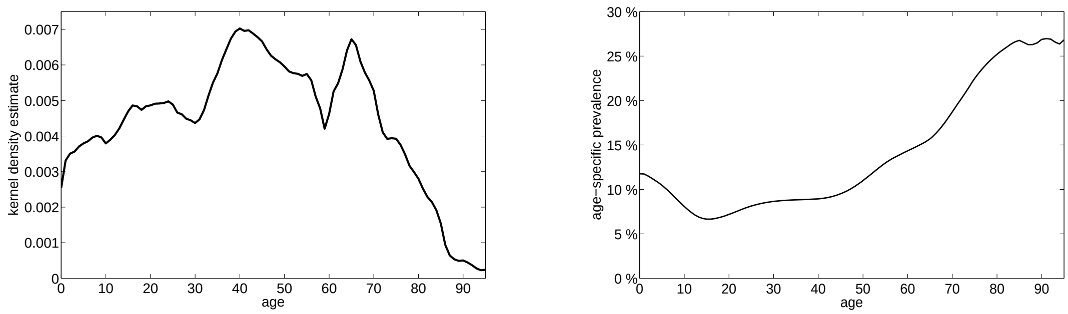
$$\hat{\hat{\beta}}(z) = \left[\sum_{i=1}^{N} k_{h_z}(u_i) w_i(z)\, X_i' X_i\right]^{-1} \times \left[\sum_{i=1}^{N} k_{h_z}(u_i) w_i(z)\, X_i' \tilde{y}_i\right]. \tag{14}$$

Note that $X_i = (1\ r_i(z))$ depends on z, as the local fractional rank from equation (13) has been used for all computations (for simplicity, we write X_i in place of $X_i(z)$ here). As a kernel function, the quartic kernel $K(u_i) = (1 - u_i^2)^2\ \mathcal{I}\,(|u_i| < 1)$ with $K_{h_z}(\cdot) = \frac{1}{h_z} K\left(\frac{\cdot}{h_z}\right)$ and $\|K_2^2\| = \int_{-\infty}^{\infty} K^2(u)\, du = \frac{256}{315}$ was chosen. Owing to the tradeoff between bias and uncertainty, the bandwidth was locally chosen as $h_z = \frac{\hat{\sigma}_z}{\left(N\hat{f}(z)\right)^{0.2}}$, where $\hat{f}(z)$ is the estimated kernel density at a particular value of z and $\hat{\sigma}_z$ is the standard deviation of z obtained from the sample.

As Kakwani et al. (1997) and Wagstaff and van Doorslaer (2000) pointed out, the error term ϵ is likely to exhibit autocorrelation. The former present an estimator for purely random samples without application of sample weights, however, this approach seems inapplicable here. Wildman (2003) proposes computation of heteroscedasticity and autocorrelation consistent standard errors as presented by Newey and West (1987). The proposal by White (1980) can be generalized as

$$\Omega_{hac}(z) = \left[f\,(z)\ E\,(X'\,X\,|\,z)\right]^{-1} \times \left[\Gamma_{hac}(z)\right] \times \left[f\,(z)\ E\,(X'\,X\,|\,z)\right]^{-1}. \tag{15}$$

where $E(X'X\,|\,z)$ and the kernel density are computed as above. The Nadaraya-Watson estimator is applied to $\Gamma_{hac}(z)$ by defining $\Psi_j(z) =$



(a) (b)

Figure 1: Empirical density of age (a) and smoothed age specific prevalence of sickness within the preceding four weeks (b)

$\sum_{i=j+1}^{N} k_{h_z}(u_i) w_i(z)\, \epsilon_i \epsilon_{i-j} \left(x_i x_{i-j}' + x_{i-j} x_i'\right)$ and $\Psi_0 = \sum_{i=1}^{N} k_{h_z}(u_i) w_i(z)\, \epsilon_i^2 x_i' x_i$,

$$\hat{\Gamma}_{hac}(z) = \hat{f}(z) \left(\Psi_0(z) + \sum_{j=1}^{m} \omega_{j,m}\, \Psi_j(z)\right) \left\|K_2^2\right\|.$$

Bartlett weights $\omega_{j,m} = 1 - \frac{j}{m+1}$ are applied to assure a positive semi-definite covariance matrix (Newey and West, 1987).

7 Empirical Results

Figure 1a describes the kernel density estimate $\hat{f}(z)$ with respect to age and corresponds with the population pyramid for Germany. Without adjusting for mortality, the cohorts born after the 1960s are smaller than those born earlier. One may see this as evidence for an aging society (see e.g. von Weizsäcker, 1996).

Figure 1b presents the smoothed age specific prevalence of illness within the preceding four weeks. The graph suggests that children younger than 10 have a higher prevalence than individuals aged between 10 and 40 years. From 45 years onwards, prevalence increases almost linearly with a short stagnation around 60 to 65.

Computing the concentration index C for the four weeks prevalence of illness yields $\hat{C} = -0.0616$. The δ method standard error of $\hat{C}$ is $\hat{\sigma}_C = 0.002$; the 95 percent confidence interval for the concentration index is $(-0.0655; -0.0576)$.

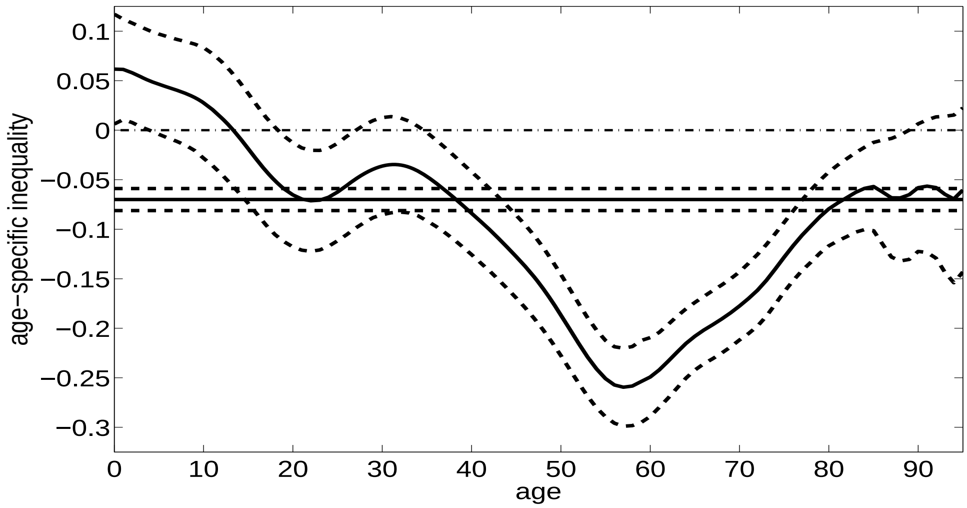


Figure 2: Wagstaff index (solid straight line) and varying inequality index (solid line) with 95 percent δ method confidence interval (dashed lines)

The corresponding Wagstaff index is $\hat{W} = -0.07$ with a standard error $\hat{\sigma}_W = 0.0057$. This yields a 95 percent confidence interval of $(-0.0811; -0.0589)$. The negative and highly significant concentration and Wagstaff indices suggest a linear bias in the health distribution to the detriment of the poor.

The varying inequality index in figure 2 varies around the overall sample estimate for the Wagstaff index. The graph suggests a statistically significant pro rich bias in illness for young children that becomes insignificant around the age of five. Except for a couple of years in the early twenties, we do not observe a significant pro poor bias for those younger than thirty. After increasing almost linearly, the strongest inequality occurs for people aged 57 where the varying index exceeds -0.25. Note that the corrected index accounts for the inverse relation between the (absolute) theoretical bounds of the concentration index and the prevalence. One may therefore consider an index of -0.25 as 25 percent of maximum possible pro poor bias given the age specific prevalence of approximately 15 percent. We find no significant inequality for people older than 80. One may consider this as support for the age as leveler hypothesis (Case and Deaton, 2005; Deaton and Paxson, 1998; Elo and Preston, 1996; Herd, 2006; Kim and Miech, 2009; Kunst and Mackenbach, 1994).

8 Concluding remarks

In this article, we combine the notion of concentration indices (Erreygers, 2009; Kakwani et al., 1997; van Kippersluis et al., 2009; Wagstaff et al., 1991; Wagstaff, 2005) with semiparametric regression techniques (Hastie and Tibshirani, 1993; Li et al., 2002) to a semiparametric inequality index with some convenient properties. Coping without a priori stratification, it adapts itself to the data, say, in terms of age or income distributions. It is by no means restricted to age as smoothing parameter. Possible modifications include varying concentration or Gini indices of inequalities for cardinal measures using the correction formula derived by Erreygers (2009) and applied by van Kippersluis et al. (2009).

Using German microcensus data, we demonstrate the power of our semiparametric approach to describe age specific income related inequalities. As a main result, we find that direction and extent of the income related bias varies considerably with age. While children exhibit pro rich inequality, strong inequalities to the detriment of the poor are observed for people aged between 30 and 80. Alike van Kippersluis et al. (2009), we observe the strongest inequality around the common age for retirement; which one may consider as support for the age as leveler hypothesis (see e.g. Case and Deaton, 2005 or Deaton and Paxson, 1998). Using cross sectional data, however, involves some limitations. Beckett (2000) points out that leveling may be an artificial effect owing to mortality selection; and Prus (2004) argues that one would require panel data to test the accumulation hypothesis. In line with van Kippersluis et al. (2009) and van Kippersluis et al. (2010), our aim was not to test the causal impact of socioeconomic status on health but to illustrate the variation of disparities in health across age groups in Germany; and thereby to provide a consistent method to measure age specific inequalities.

References

Balia, S., Jones, A. M., 2008. Mortality, lifestyle and socio-economic status. J Health Econ 27 (1), 1–26.

Banks, J., Marmot, M., Oldfield, Z., Smith, J. P., Jan. 2007. The SES health gradient on both sides of the Atlantic. IZA Discussion Papers 2539, Institute for the Study of Labor (IZA).

Beckett, M., Mar 2000. Converging health inequalities in later life–an artifact of mortality selection. J Health Soc Behav 41 (1), 106–119.

Case, A. C., Deaton, A., 2005. Broken down by work and sex: how our health declines. In: Wise, D. A. (Ed.), Analyses in the Economics of Aging. Chicago: Chicago University Press, Ch. 6, pp. 185–205.

Deaton, A. S., Paxson, C. H., May 1998. Aging and inequality in income and health. American Economic Review 88 (2), 248–53.

Elo, I. T., Preston, S. H., Jan 1996. Educational differentials in mortality: United States, 1979-85. Soc Sci Med 42 (1), 47–57.

Erreygers, G., Mar 2009. Correcting the concentration index. J Health Econ 28 (2), 504–515.

FSO, 2006. Lebenslagen in Deutschland. Federal Statistical Office.

Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. Journal of the Royal Statistical Society. Series B (Methodological) 55 (4), 757–796.

Herd, P., 2006. Do functional health inequalities decrease in old age?: Educational status and functional decline among the 1931-1941 birth cohort. Research on Aging 28 (3), 375–392.

Humphries, K. H., van Doorslaer, E., 2000. Income-related health inequality in Canada. Soc Sci Med 50 (5), 663–671.

Kakwani, N., Wagstaff, A., van Doorslaer, E., 1997. Socioeconomic inequalities in health: Measurement, computation, and statistical inference. Journal of Econometrics 77 (1), 87–103.

Kim, J., Durden, E., Dec 2007. Socioeconomic status and age trajectories of health. Soc Sci Med 65 (12), 2489–2502.

Kim, J., Miech, R., Feb 2009. The black-white difference in age trajectories of functional health over the life course. Soc Sci Med 68 (4), 717–725.

Kunst, A. E., Mackenbach, J. P., Jun 1994. The size of mortality differences associated with educational level in nine industrialized countries. Am J Public Health 84 (6), 932–937.

Lechert, Y., Schimpl-Neimanns, B., 2007. GESIS-ZUMA-Methodenbericht 2007/08 Mikrozensus Scientific Use File 2005 Dokumentation und Datenaufbereitung. Federal Statistical Office.

Lerman, R. I., Yitzhaki, S., 1989. Improving the accuracy of estimates of Gini coefficients. Journal of Econometrics 42 (1), 43–47.

Li, Q., Huang, C. J., Li, D., Fu, T.-T., 2002. Semiparametric smooth coefficient models. Journal of Business & Economic Statistics 20 (3), 412–22.

Lynch, S. M., May 2003. Cohort and life-course patterns in the relationship between education and health: a hierarchical approach. Demography 40 (2), 309–331.

Mirowsky, J., Ross, C. E., 2005. Education, cumulative advantage, and health. Ageing International 30, 27–62.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.

Prus, S. G., 2004. A life course perspective on the relationship between socioeconomic status and health: testing the divergence hypothesis. Can J Aging 23 Suppl 1, S145–S153.

Rao, C., 1965. Linear Statistical Inference and its Applications. New York: Wiley.

Reeske, A., Spallek, J., Razum, O., 2009. Changes in smoking prevalence among first- and second-generation Turkish migrants in Germany - an analysis of the 2005 Microcensus. International Journal for Equity in Health 8, 26–34.

Ross, C. E., Wu, C. L., Mar 1996. Education, age, and the cumulative advantage in health. J Health Soc Behav 37 (1), 104–120.

van Doorslaer, E., Koolman, X., Jul 2004. Explaining the differences in income-related health inequalities across European countries. Health Econ 13 (7), 609–628.

van Doorslaer, E., Koolman, X., Jones, A. M., Jul 2004. Explaining income-related inequalities in doctor utilisation in Europe. Health Econ 13 (7), 629–647.

van Doorslaer, E., Wagstaff, A., Bleichrodt, H., Calonge, S., Gerdtham, U. G., Gerfin, M., Geurts, J., Gross, L., Häkkinen, U., Leu, R. E., O'Donnell, O., Propper, C., Puffer, F., Rodríguez, M., Sundberg, G., Winkelhake, O., 1997. Income-related inequalities in health: some international comparisons. J Health Econ 16 (1), 93–112.

van Kippersluis, H., O'Donnell, O., van Doorslaer, E., Ourti, T. V., Feb 2010. Socioeconomic differences in health over the life cycle in an egalitarian country. Soc Sci Med 70 (3), 428–438.

van Kippersluis, H., Ourti, T. V., O'Donnell, O., van Doorslaer, E., Jul 2009. Health and income across the life cycle and generations in Europe. J Health Econ 28 (4), 818–830.

von Weizsäcker, R. K., 1996. Distributive implications of an aging society. European Economic Review 40 (3-5), 729 – 746, papers and Proceedings of the Tenth Annual Congress of the European Economic Association.

Wagstaff, A., 2005. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. Health Econ 14 (4), 429–432.

Wagstaff, A., Paci, P., van Doorslaer, E., 1991. On the measurement of inequalities in health. Soc Sci Med 33 (5), 545–557.

Wagstaff, A., van Doorslaer, E., 2000. Measuring and testing for inequity in the delivery of health care. The Journal of Human Resources 35 (4), 716–733.

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 (1), 207 – 223.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48 (4), 817–838.

Wildman, J., 2003. Income related inequalities in mental health in Great Britain: analysing the causes of health inequality over time. J Health Econ 22 (2), 295–312.

Willson, A. E., Shuey, K. M., Elder, Jr., G. H., 2007. Cumulative advantage orocesses as mechanisms of inequality in life course health. American Journal of Sociology 112 (6), 1886–1924.