Measuring variations in health inequalities: Semiparametric modeling of the concentration index

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2012

vorgelegt

von
Diplom-Volkswirt Martin Siegel
aus
Gießen

Referat: Professor Dr. Karl Mosler
Korreferat: Professor Dr. Markus Lüngen
Tag der Promotion: 14. Dezember 2012

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> *"In any discussion of social equity and justice, illness and health must figure as a major concern.* [...] *First, health is among the most important conditions of human life and a critically significant constituent of human capabilities which we have reason to value. Any conception of social justice that accepts the need for a fair distribution as well as efficient formation of human capabilities cannot ignore the role of health in human life and the opportunities that persons, respectively, have to achieve good health - free from escapable illness, avoidable afflictions and premature mortality."* Amartya Sen[1]

The relevance of equity in health for a fair and equitable development of human resources and societies, respectively, has been well explained by Sen (2002). Individual incomes, social arrangements, genetic propensities, working conditions and the epidemiological environments are among the factors which may contribute to health achievements or failures. In his article, Sen (2002) points out that discussions on equity in health should by no means be restricted to the question how health care is distributed. Although Sen stresses that health inequity is not equivalent to health inequality and that concentrating on the latter is not sufficient to assess

[1] Amartya Sen: Why Health Equity? Keynote Address to the Third Conference of the International Health Economics Association; York, 23 July 2001 (published in *Health Economics* as Sen, 2002)

the former, he still agrees that inequalities in health are a matter of interest of its own. "*It* (health inequality) *does have interest of its own, and it certainly is a very important part of our understanding of health equity, which is a broader notion*" (Sen, 2002, p. 662). Needless to say that equity should be considered as a concept which is probably neither objective nor directly measurable, but rather a social and political consensus to be found. The present thesis tries to contribute to this process by extending the empirical basis. More precisely, it is focused on the introduction and application of a new econometric approach to measure variations in health inequalities.

The concentration index is an adaptive tool for the measurement of income-related inequalities which offers some advantages over techniques such as, say, comparisons of prevalences in income quintiles or other population subgroups. In order to explain the causes of health sector inequalities, 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 concentration index of the health variable as the sum of the concentration indices of the explanatory variables weighted by their respective elasticities. From this decomposition approach, one may infer how health inequality would change if, say, no demographic effects were present.[2] Jones and López Nicolás (2006) further separate the contribution of response heterogeneity from the unexplained (residual) part in the decomposition formula derived by Wagstaff et al. (2003). They point out that response heterogeneity in the elasticities may change the composition of health gradients considerably.[3] During a research project on health and health care utilization inequalities in Germany (Lüngen et al., 2009), however, a major shortcoming of both decomposition approaches became evident: They do not allow comparisons of inequalities across age groups. When trying to gain a better understanding of the origins of health inequalities in Germany, the question in

[2]Note that this method has extensively been applied in Lüngen et al. (2009).
[3]More precisely, Jones and López Nicolás (2006) analyze heterogeneity in terms of age and sex.

which age groups inequalities could actually be observed came up repeatedly. At the time the research described in Lüngen et al. (2009) was finished, there seemed to be no suitable method to measure age-specific income-related health inequalities.

This lack of suitable methods to describe variations in health inequalities is the point of departure for the present thesis. Chapter 2 presents a new econometric approach to measure variations in inequalities with respect to some other continuous or discrete explanatory variable. The proposed model combines the well-established concentration index with the smooth coefficient model introduced by Li et al. (2002) to a semiparametric inequality index. The chapter introduces the model and presents a feasible estimator which is applied to data on self-assessed health from the 2005 and 2009 surveys of the German microcensus. It is found that the extent of health inequalities varies considerably between age groups. While no inequalities to the detriment of the poor are observed among children and the elderly, considerable health disadvantages are found among the economically deprived particularly in later mid-life. The method is compared with the results obtained when estimating concentration indices for five year interval age groups. The estimator's behavior in smaller samples is assessed through drawing subsamples from the data. The results suggest that the index works well for normal and large sized samples with more than thirty to forty thousand observations.

One may argue that the health variable in chapter 2 is rather general and subjective. Chapter 3 uses data on specific diseases drawn from the TNS Health Care Access Panel. Using the methods introduced in chapter 2, age-specific variations of income inequalities and income-related health inequalities in obesity, hypertension and diabetes are addressed. The results show that the social gradients concerning the prevalence of the diseases under consideration vary considerably with age in the female sample. Interestingly, only marginal variations are found in the male sample. Similar to chapter 2, the strongest inequalities to the detriment of the economically deprived are found in later mid-life.

While chapters 2 and 3 illustrate health outcomes, chapter 4 addresses smoking as an addictive behavior being a major threat to one's health. Sen (2002) stresses the distinction between a free decision not to care for one's health and failure to achieve good health because of inadequate social arrangements and lacking resources. He points out that "*indeed even smoking and other addictive behaviour can also be seen in terms of a generated 'unfreedom' to conquer the habit*" Sen (2002, p. 660). The method from chapter 2 is applied to data from the 2005 and 2009 surveys of the German microcensus. Chapter 4 addresses the prevalence of current and ever-smoking as well as smoking cessation. The result suggest that men are, on average, more likely to be current smokers and ever-smokers than women. Current and ever-smoking are both concentrated in worse-off households in younger birth cohorts. Smoking cessation is concentrated among the better-off in all age groups. The results from the 2005 and 2009 samples are quite similar. This may be seen as a hint that the microcensus data on smoking are fairly reliable.[4]

Although age-specific variations in health inequalities have been in the main focus in chapters 2, 3 and 4, the varying inequality index is by no means restricted to age. The factor by which inequalities are allowed to vary may be replaced by any meaningful variable. As an example, the Index of Multiple Deprivation used by Maier et al. (2011) is employed in chapter 5 as both, the socioeconomic status variable and in place of age. The present thesis goes beyond a pure application of the German Index of Multiple Deprivation and uses the within-domain rankings to propose an empirically driven weighting scheme based on factor analysis. Comparing the theory driven and the empirically driven approaches, it is found that the weighting scheme derived from factor analysis puts considerably more emphasis on income and unemployment rates than the index derived from the literature. Both deprivation indices are highly correlated and produce similar rankings, though. As in chapter 3, data on obesity, hypertension, diabetes, income, age and sex are drawn from the TNS Health Care Access Panel. The

[4]Note that no household should have participated in both surveys as they are only included in four subsequent waves and are removed then. Hence, the last households included in 2005 were removed in 2008 and the samples are non-overlapping. The comment that the microcensus data are fairly reliable is motivated by the fact that the two non-overlapping samples produce almost equivalent results.

results for age-specific community deprivation-related health inequalities are similar to those found for age-specific income-related inequalities. The results for the deprivation-specific variations of income-related health inequalities, however, yield no significant variations. It is found in chapter 5 that the results obtained when using the theory driven and the empirically driven deprivation indices look rather similar. When measured for communes with similar deprivation status, income-related health inequalities are less marked than the overall-sample income-related inequalities. They do, however, not vary with the community deprivation rank.

In summary, this thesis presents evidence that income-related inequalities to the detriment of the economically deprived households exist in Germany. While this holds for health status in almost all age groups, smoking only concentrates among the worse-off households in younger cohorts. Considering the results for ever-smoking as a cohort effect, the results suggest that smoking indeed changed from a pro-rich towards a pro-poor habit over the twentieth century. While individual income and community deprivation are similarly associated to individual health, income-related health inequalities do not vary with community deprivation.

Chapter 2

Semiparametric modeling of age-specific variations in income-related health inequalities[1]

2.1 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; van Doorslaer et al., 1997; van Doorslaer and Koolman, 2004; van Doorslaer et al., 2004; Jones and López Nicolás, 2006; Wagstaff et al., 1991; Wagstaff and van Doorslaer, 2000; Wagstaff et al., 2003). 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., 2009, 2010). 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

[1]This chapter is based on a manuscript co-authored by Karl Mosler.

Britain (Banks et al., 2007) and the Netherlands (van Kippersluis et al., 2010).

The literature has long been divided into two competing hypotheses explaining why disparities in health may differ over the life course: the disadvantage accumulation hypothesis and the age as leveler hypothesis (Dupre, 2007). Both hypotheses agree that low socioeconomic status and lacking resources are associated with less healthy lifestyles, higher health risks, a faster decline of health status and higher mortality rates. The disadvantage accumulation hypothesis argues that social gradients in health develop in early life and become stronger as socioeconomic and health disadvantages accumulate over the entire life course (Kim and Durden, 2007; van Kippersluis et al., 2010; Lynch, 2003; Mirowsky and Ross, 2005; Prus, 2004; Ross and Wu, 1996; Willson et al., 2007). The age as leveler hypothesis adds the assumption that the decline of health is an unavoidable part of aging. Followers of the age as leveler hypothesis argue that health inequalities evolving from socioeconomic disadvantages increase up to some point in midlife but are then outperformed by the decline in health status due to aging and to some extent attenuated by a relief through retirement (Case and Deaton, 2005; Deaton and Paxson, 1998; Elo and Preston, 1996; Kim and Miech, 2009; van Kippersluis et al., 2010; Kunst and Mackenbach, 1994). The two hypotheses have long been treated as competing. However, van Kippersluis et al. (2010) argue that evidence found for the age as leveler hypothesis is also consistent with the disadvantage accumulation hypothesis. Some authors stress that leveling may be an artificial result due to mortality selection, though Beckett (2000) points out that this needs not necessarily be true for self-reported health.

A common approach to measure health inequalities is the concentration index. Although it requires some ranking, say, by income, its computation does not require predefined socioeconomic 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, however, their results favor the age as leveler hypothesis for France, Germany and the U.K.

This chapter introduces a varying inequality index for dichotomous health variables which does not require a priori sample stratification. As in van Kippersluis et al. (2009), the objective of this chapter is to describe the differences in income-related health inequalities across age cohorts. One may consider estimating a nonparametric smoother through age-specific concentration indices as an alternative, however, this approach would reduce the reliability of the results considerably. The number of observations in predefined age groups of, say, five year intervals may be rather small resulting in high uncertainty particularly among the oldest age groups. Further, the estimates for ages close to the upper (lower) cohort limits would only be subject to the younger (older) individuals within the same cohort and hence likely be biased. Smoothing over such results in a second step would then add its own uncertainty and lead to fairly imprecise results. Based on the varying coefficient model (Hastie and Tibshirani, 1993; Li et al., 2002), a semiparametric extension of the convenient regression approach (Kakwani et al., 1997) is proposed here. Using kernel smoothing techniques and a locally chosen bandwidth allows to estimate the functional relationship between the concentration index and age. The varying index is adjusted using Wagstaff's (2005) correction formula for binary variables with local estimates of the mean.

2.2 Methods

2.2.1 The concentration index

In their paper, Wagstaff et al. (1991) introduce the concept of concentration curves and indices to the field of health inequality analyses. The concentration index C stems from the concentration curve where the cumulative share of some health variable y is plotted against the cumulative share of the population ranked by socioeconomic position (Wagstaff et al., 1991). It measures twice the area between the concentration curve and the line of equality. C is bound in the $(-1;1)$ interval and becomes positive (negative), if the variable of interest concentrates among the rich (poor). C is zero if no income-related inequality is observed.

Using the covariance approach in Lerman and Yitzhaki (1989), Kakwani et al. (1997) present the regression formula for the concentration index

$$\frac{2\sigma_r^2}{\mu} y = \beta_0 + \beta_1 r + \varepsilon, \tag{2.1}$$

where μ is the mean of the health variable and r is the fractional rank with variance σ_r^2. Equation (2.1) can be estimated using linear regression models to obtain the concentration index as $C = \beta_1$.

2.2.2 Varying coefficient models

In the framework of varying coefficient models, Li et al. (2002) propose a semiparametric 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 vector $(\beta_0, \ldots, \beta_Q)$ are modeled as smooth functions of another regressor z which varies in some $Z \subset \mathbb{R}$:

$$y = \beta_0(z) + \sum_{q=1}^{Q} \beta_q(z) x_q + \varepsilon. \tag{2.2}$$

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

$$\beta(z) = \left(E(X'X \mid z)\right)^{-1} E(X'y \mid z), \tag{2.3}$$

where $X = (1\ x_1 \dots x_Q)$. Li et al. (2002) have shown that, for increasing numbers of observations n, the estimator $\hat{\beta}(z)$ obtained from (2.3) asymptotically follows a normal distribution, i.e. $\sqrt{nh_z}\left(\hat{\beta}(z) - \beta(z)\right) \sim N(0, \Omega(z))$; see section 2.2.5 for the estimation of the covariance matrix.

2.2.3 A semiparametric inequality index

Combining the weighted regression approach from equation (2.1) with the varying coefficient model (2.2), the proposal for a semiparametric convenient regression formula is

$$2\,\frac{\sigma_r^2(z)}{\mu(z)}\,y = \beta_0(z) + \beta_1(z)\,r(z) + \varepsilon, \tag{2.4}$$

such that $C(z) = \beta_1(z)$, $z \in Z$. Note that the concentration index is a bivariate extension of the Gini index; if y is the social status variable, equation (2.4) works as a semiparametric Gini index. The local mean $\mu(z)$ can be estimated nonparametrically. The weighted fractional rank $r(z)$ has to be written as a function of z. 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 regardless of their individual values of z would be misleading. Technically, the condition that the sum of the sample weights has to equal 1 and the mean and variance of the weighted fractional rank have to be 0.5 and $1/12$, respectively, must hold (Lerman and Yitzhaki, 1989). This can only be fulfilled if the weighted fractional rank is computed using only those individuals included in the local regression and incorporating the kernel weights $k_{h_z}(u_i)$:

$$r_i(z) = \sum_{j=1}^{i} w_j(z) k_{h_z}(u_j) - \frac{w_i(z) k_{h_z}(u_i)}{2}. \tag{2.5}$$

The vector of sample weights $w(z)$ must be rescaled such that $\sum_{i=1}^{n} w_i(z)\, k_{h_z}(u_i) = 1$ for each $z \in Z$. Note that the mean and variance of $r(z)$ are then (asymptotically) sample independent and do not vary with z.

For binary variables, the bounds of the concentration index depend inversely on the variable's mean, $|C| \leq 1 - \mu$ (see Wagstaff, 2005, 2011; Erreygers, 2009). For an intuitive explanation, first assume a constant equal to 1. With no difference between individuals, concentration among rich or poor is impossible; the concentration index equals zero. Now consider, say, 10 percent ones and 90 percent zeros. Ordering the variable by itself, one would obtain a Gini index of 0.9; the largest possible concentration. Analogously, the maximum possible inequality for a binary variable with 90 percent ones and 10 percent zeros would equal 0.1 (see Wagstaff, 2011, for a graphical illustration). Comparisons of concentration indices of binary variables with rather different means may thus be misleading.

There is an ongoing discussion on possible correction methods for concentration indices of limited variables (Erreygers, 2009; Wagstaff, 2005, 2011) with a dissent between the authors on how a corrected index should react to changes of the mean. Considering the above example, Erreygers (2009) would argue that an increase from 10 to 20 percent implies a decrease in inequality as now the second richest (poorest) decile is also affected. The argument in Wagstaff (2005, 2011) is that, as still only the richest (poorest) are affected, the new situation still corresponds with the maximum possible inequality. A reaction of the index to pure prevalence changes would not be desirable here as the motivation of this chapter is to compare inequalities across age groups and sexes. The proposal is to adapt the formula in Wagstaff (2005, 2011) as a pointwise correction of the semiparametric concentration index using the local mean $\mu(z)$ of y:

$$W(z) = \frac{C(z)}{1 - \mu(z)}. \tag{2.6}$$

2.2.4 Estimation

Applying a consistent Nadaraya-Watson estimator, sample weights are taken into account and $\hat{C}(z) = \hat{\beta}_1(z), z \in Z$, is obtained by computing

$$\hat{\beta}(z) = \left[\sum_{i=1}^{n} k_{h_z}(u_i) w_i(z)\, X_i' X_i\right]^{-1} \left[\sum_{i=1}^{n} k_{h_z}(u_i) w_i(z)\, X_i' \tilde{y}_i\right]. \tag{2.7}$$

Note that $\tilde{y}_i = \left({}^{2\sigma_r^2(z)}/_{\mu(z)}\right) y_i$ and $X_i = (1 \quad r_i(z))$ depend on z as the local mean and the local fractional rank from equation (2.5) are involved (for simplicity, X_i is written in place of $X_i(z)$ here). The kernel weights are $k_{h_z}(u_i) = K_{h_z}(u_i) \left[\sum_{j=1}^{n} K_{h_z}(u_j)\right]^{-1}$ with $u_i = z_i - z$. The quartic kernel $K(u_i) = ({}^{15}/_{16}) \left(1 - u_i^2\right)^2 I_{|u_i|<1}$ with $\left\|K_2^2\right\| = \int_{-\infty}^{\infty} K^2(u)\, du = {}^{5}/_{7}$ is used, where I_A is an indicator function of restriction A. The bandwidth h_z is included such that $K_{h_z}(\cdot) = (h_z)^{-1} K\left({}^{\cdot}/_{h_z}\right)$. The quartic kernel assigns higher weights to observations closer to z (smaller u_i), lower weights for observations further away from z (larger u_i) and zero weights if an observation is outside the bandwidth.[2] Although the estimator is asymptotically unbiased (Li et al., 2002), any nonparametric regression in finite samples suffers to some extent from a tradeoff between bias and variability: decreasing the bandwidth parameter decreases the bias but at the cost of increasing uncertainty; and vice versa (see also Bilger, 2008, for a discussion). This problem is addressed here by choosing the bandwidth inversely to the local data density as $h_z = 1.06\, \hat{\sigma}_z\, n^{-0.2} \hat{f}_z^{-0.3}$, 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. Fan and Gijbels (1992) have shown that adaptive local smoothers generally yield good results and, in addition, avoid the well-known boundary effect.

To obtain confidence intervals for the semiparametric concentration index, its standard error needs to be estimated. Kakwani et al. (1997), Wildman (2003) and O'Donnell et al. (2008) argue that it is not sufficient to estimate the standard error of $\beta_1(z)$ from equation

[2] For $K_{h_z}(u_i)$, restriction A is $|u_i| < h_z$ making the indicator function 1 if $|u_i| < h_z$ and 0 otherwise.

(2.1) because of the sample variability of $\mu(z)$. One may estimate $\beta_0^*(z)$ and $\beta_1^*(z)$ from $y = \beta_0^*(z) + \beta_1^*(z)r(z) + \varepsilon^*$ and consider the concentration index as a nonlinear combination of the two coefficients with $\beta_0^*(z) + 0.5\beta_1^*(z)$ in place of $\mu(z)$. The variance can be approximated using the δ method (Rao, 1965) on $C(z) \approx 2\sigma_r^2(z)\left[\beta_0^*(z) + 0.5\beta_1^*(z)\right]^{-1}\beta_1^*(z)$ for the semiparametric concentration index.[3] The variance of the varying Wagstaff index $W(z)$ can be estimated analogously.[4] Further, it has been argued that the covariance matrix from a simple OLS regression is not wholly accurate because the error term ε^* may be autocorrelated and heteroscedastic (Kakwani et al., 1997). This chapter follows Wildman (2003) who proposes using the order of the rank variable in place of time to compute pointwise heteroscedasticity and autocorrelation consistent Newey-West covariance matrices (see section 2.2.5 for details).

2.2.5 Variance estimation

According to Li et al. (2002), the covariance matrix $\Omega(z)$ in the semiparametric varying coefficient model is

$$\Omega(z) = \left[f_z\, E\left(X'X \mid z\right)\right]^{-1}\ \Phi(z)\ \left[f_z\, E\left(X'X \mid z\right)\right]^{-1} \tag{2.8}$$

[3]This yields

$$\sigma_C^2(z) \approx 4\sigma_r^4(z)\frac{\beta_0^2(z)\sigma_{11}(z) + \beta_1^2(z)\sigma_{22}(z) - 2\beta_1(z)\beta_0(z)\sigma_{12}(z)}{(\beta_0(z) + 0.5\beta_1(z))^4}$$

for the variance $\sigma_C^2(z)$ of $C(z)$, where the $\sigma_{ij}(z)$ are the ijth elements of the covariance matrix $\Omega(z)$ of $\beta(z)$.

[4]Equation (2.6) for $W(z)$ can be written as

$$W(z) = \frac{C(z)}{1-\mu(z)} \equiv \frac{2\sigma_r^2(z)}{(\beta_0(z) + 0.5\beta_1(z))(1 - \beta_0(z) - 0.5\beta_1(z))}\beta_1(z).$$

One may then estimate the variance $\sigma_W^2(z)$ of $W(z)$ as

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

with $\Phi(z) = f_z \, E\left(X'X \ \sigma^2_\varepsilon(z) \mid X, z\right) \left\|K_2^2\right\|$ and $\sigma^2_\varepsilon(z) = E(\varepsilon_i^2 \,|X, z)$. To estimate a heteroscedasticity and autocorrelation consistent covariance matrix $\hat{\Omega}_{hac}(z)$, $\Phi(z)$ must be adapted accordingly. Following the proposal by White (1980), $\Phi(z)$ is computed as

$$\hat{\Phi}_{hac}(z) = f_z \ \left(\Psi_0(z) + \sum_{j=1}^{m} \omega_{j,m}\Psi_j(z)\right) \left\|K_2^2\right\|. \tag{2.9}$$

with

$$\Psi_j(z) = \sum_{i=j+1}^{n} k_{h_z}(u_i) w_i(z) \ \varepsilon_i \varepsilon_{i-j} \left(x_i x'_{i-j} + x_{i-j} x'_i\right) \tag{2.10}$$

and $\Psi_0 = \sum_{i=1}^{n} k_{h_z}(u_i) w_i(z) \varepsilon_i^2 x'_i x_i$. Bartlett weights $\omega_{j,m} = 1 - j/(m+1)$ are applied to assure a positive semi-definite covariance matrix (Newey and West, 1987), $E(X'X \,|\, z)$ and the kernel density are computed as above.

2.3 Data and variables

Data for the empirical application were drawn from the 2005 survey of the German microcensus provided by the Research Data Centers of the Federal Statistical Office and the Statistical Offices of the Federal States (Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder). 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).

The scientific use file (SUF) available for non-profit research organizations is used for the empirical illustration. The SUF comprises a randomly drawn subsample of approximately 70 percent $(n = 477,239)$ of the German microcensus. Inverse probability weights accounting for the regional, age and sex specific composition of the sample are provided by the Federal Statistical Office (see Lechert and Schimpl-Neimanns, 2007, for a technical report). Removing $92,458$ individuals from the sample owing to missing information leaves $n = 384,781$ observations ($198,877$ female and $185,904$ male) for the empirical analysis. The inverse probability weights were adjusted accordingly.

The health outcome variable is a subjective measure of health, i.e. whether an individual experienced illness including chronic diseases during the preceding four weeks. Although positive responses may include (common) acute diseases such as colds or light flues, one may assume that these affect all socioeconomic groups. The health outcome is measured 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”. The analysis is restricted to having been ill and a binary variable with outcome 1 for “yes, ill” is generated. The options “no” and “yes, injured” were both treated as “not ill” and hence coded as 0. Assuming that only those who actually felt affected would consider themselves as ill, this measure of health has the advantage that it only includes diseases if they were relevant to the respective individual.

Household income is used to assess an individual’s relative socioeconomic position (note that this is not restricted to a particular source of income in the data). One may consider this as unsuitable for some countries particularly after retirement. However, Germany is an exemption because its welfare policies can be seen as rather status preserving (Brockmann et al., 2009). Approximately 90 percent of the German population are covered by the public pension system where benefits after retirement depend on compulsory contributions subtracted from

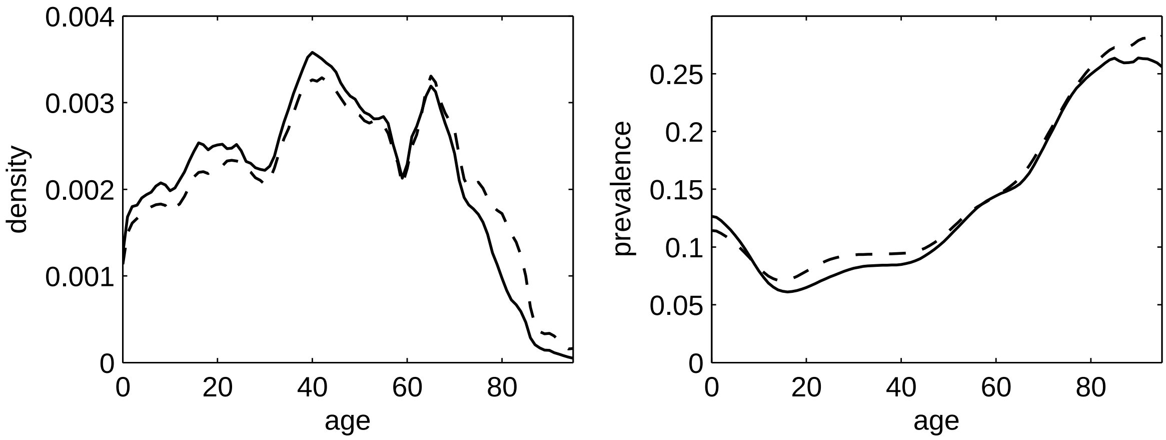


Figure 2.1: Descriptive figures (2005 sample)
Empirical density of age (left) and smoothed age-specific prevalence of sickness within the preceding four weeks (right) for males (solid lines) and females (dashed lines).

the gross income (for a description of the German public pension system, see Boersch-Supan and Wilke, 2004). Therefore, considerable changes of the relative socioeconomic position within one's age group after retirement seem unlikely and household income can be seen as a suitable indicator for the socioeconomic position over the entire life course. The modified OECD equivalence scale is used 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).

2.4 Results

2.4.1 Results from the 2005 sample

The left graph in figure 2.1 describes the kernel density estimate $\hat{f}_z$ of the nonparametric smoothing regressor z (age). The graphs for the male and female samples imply that the largest bandwidth parameter was used for subjects older than 80 for both sexes. The distribution of age groups corresponds with the population pyramid for Germany. Without adjusting for mortality, birth cohorts younger than 40 (i.e. born after 1965) are smaller than those born earlier. One may see this as evidence for an aging society (see e.g. von Weizsäcker, 1996).

The right graph in figure 2.1 presents the smoothed age-specific prevalence of illness within the preceding four weeks. The graphs suggest that children younger than 10 have a higher prevalence than individuals aged between 10 and 40 years. From 45 years onward, prevalence increases almost linearly and stagnates for the elderly older than 80. While the prevalence is somewhat lower for female children, the graph suggests that it is slightly higher among female adolescents and adults until around 45 and for elder women over 80.

The upper and lower left graphs in figure 2.2 present the age-specific means of the net equivalent household incomes for males and females, respectively. Considering that income is assigned equally to each household member, the graphs suggest that households with children have, on average, the lowest net equivalent household income. While the age-specific mean income does not differ between sexes for children, adult men have, on average, higher incomes. The sex difference appears to be highest for the oldest (over 75). Between the age of 20 and 60, one may say that the older an individual the higher is the expected net equivalent household income. The bump around age 40 for in both graphs stems from the higher average number of dependent children which increases the equivalence weights in the corresponding households. The expected mean income peaks around 57, decreases with retirement (i.e. for subjects aged between 57 and 70) and varies around $1,400$ Euro for males and $1,200$ Euro for females older than 70.

The homogeneous (age-independent) Gini index is 0.282 with a standard error of 0.0033 for the female and 0.2915 with a standard error of 0.0036 for the male sample. The right hand graphs in figure 2.2 present the corresponding age-specific Gini indices. The age-specific index varies around 0.27 for children and adolescents in both samples. For males over 20, the graph suggests an increasing income inequality which peaks at age 57 with a Gini of 0.32. In

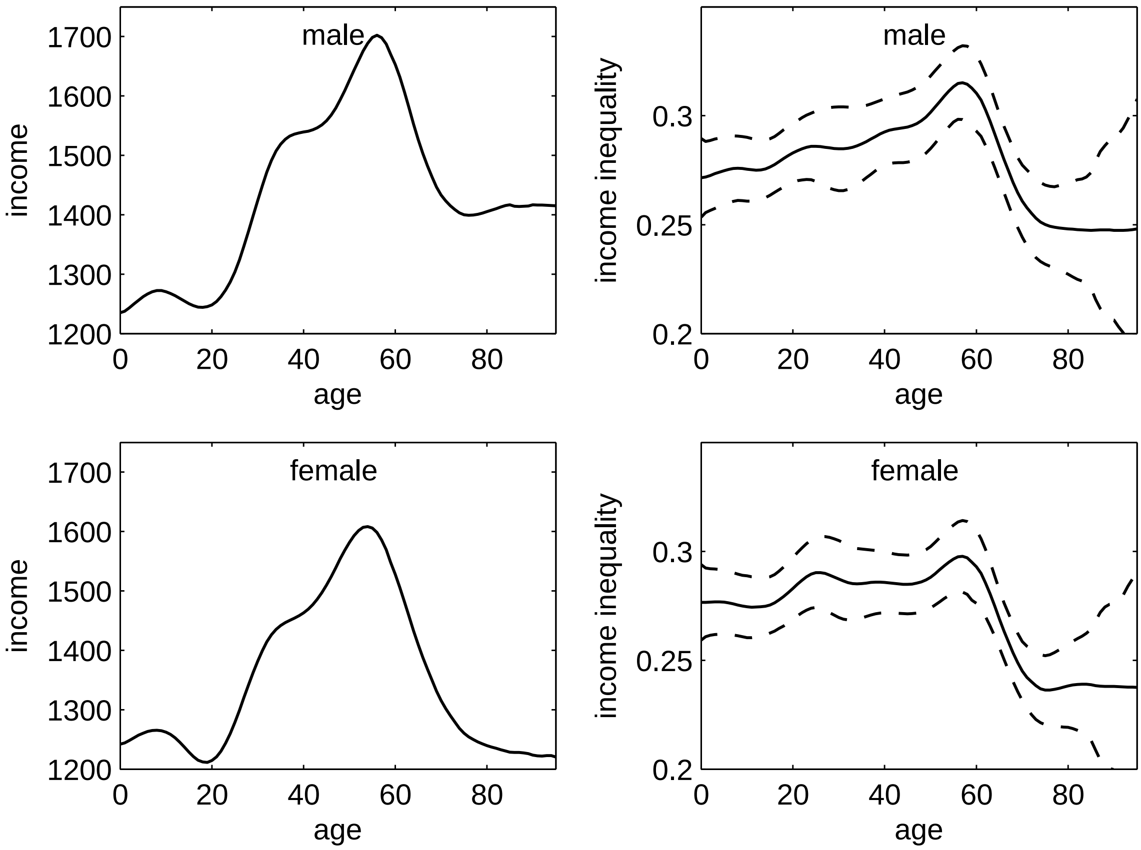


Figure 2.2: Age-specific income and Gini indices (2005 sample)
Age-specific mean (left) and Gini indices (right) with 95 percent confidence intervals (dashed lines) of net equivalent household income for males (top) and females (bottom).

the female sample, income inequality is somewhat higher among adults (Gini of 0.28) than among children. It increases between 50 and 60 and peaks at age 58 with a Gini of 0.3. With the retirement age, the index drops back to approximately 0.25 for both sexes.

Computing the homogeneous concentration indices for the four weeks prevalence of illness yields -0.0606 with a standard error of 0.0026 for the female and -0.0653 with a standard error of 0.0028 for the male sample. The homogeneous Wagstaff indices are -0.0696 with a standard error of 0.0074 for the female and -0.0738 with a standard error of 0.0079 for the male sample. The negative and highly significant concentration and Wagstaff indices suggest

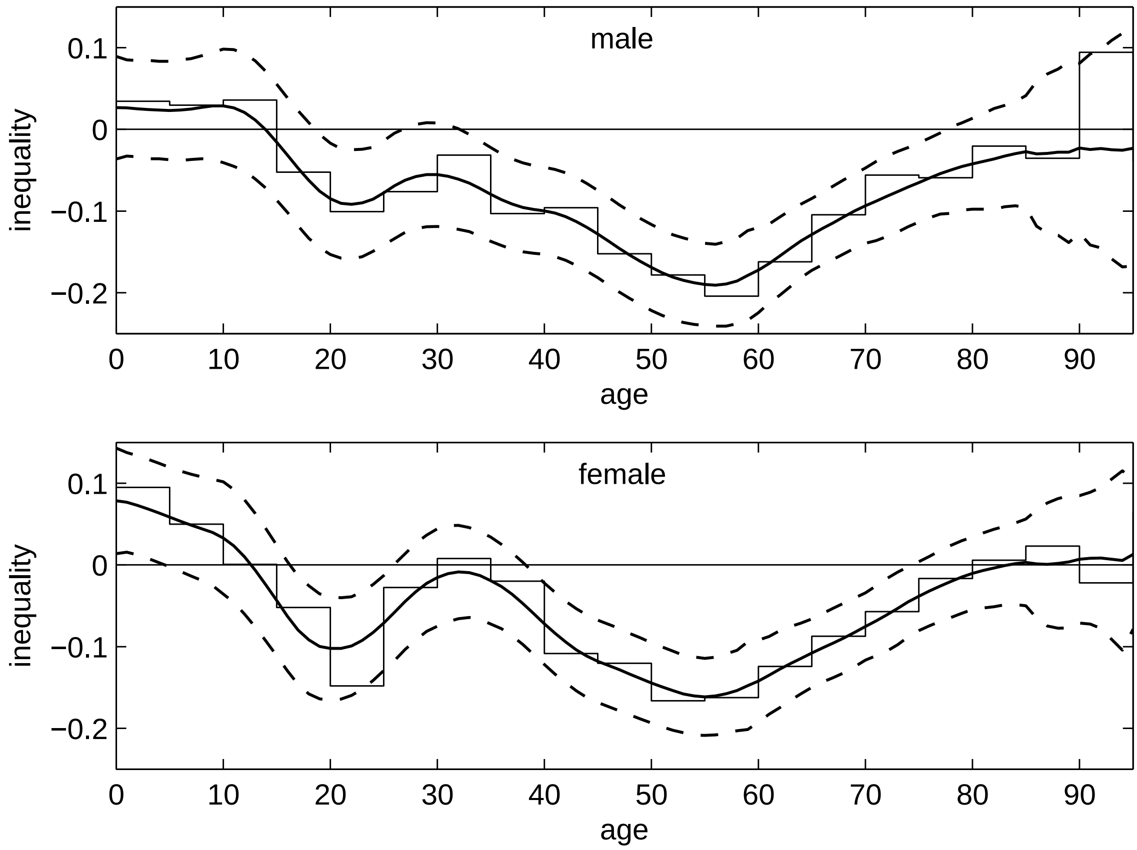


Figure 2.3: Varying inequality index (2005 sample)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for males (top) and females (bottom); the step function represents concentration indices computed for five year age groups.

a considerable concentration of illness among the poor.

The varying Wagstaff indices in figure 2.3 vary around the homogeneous estimates. The graph for the male sample suggests a statistically insignificant concentration of illness among the better-off for children younger than 14. Illness is significantly concentrated among males in lower income households in the age groups between 19 and 26 as well as between 33 and 77. There is no significant concentration among elder males over 77. The graph for the female sample suggests a significant concentration of illness among higher income households for children younger than 5. As in the male sample, the index for females shifts from a concentration among the better-off towards a concentration among the worse-off in late childhood. Significant concentration among females in lower income households is found for those between 17 and 25 as well as between 39 and 74. No significant concentration is found for the elderly over 74. Considerable sex-specific differences in the curve shapes are only found for the 20 to 40 years old. While there is some flattening for the male sample, inequality is lower among females in this age group than around the 20 as well as the 40 to 70 years old.

Comparing the results of the varying and the homogeneous Wagstaff indices, it is observed that the homogeneous index significantly underestimates the concentration of illness among the lower income households for females between 44 and 65 and males between 44 and 67. Conversely, the homogeneous indices would suggest a significant concentration of illness among children in lower income households while the age-specific indices yield opposite or insignificant results. The results from the index computed for five year age groups are somewhat similar to those from the varying index where data are dense. However, the five year interval index exhibits several leaps and a considerably higher variability. Comparison of the two graphs demonstrates the above mentioned bias close to the group limits.

2.4.2 Results from the 2009 sample[5]

The patterns of the kernel density estimates and the age-specific prevalences in figure 2.4 are similar to those in figure 2.1. It should be mentioned that the level of the kernel density estimate differs somewhat and that the estimated prevalence rates are higher across all age groups in the 2009 sample than in the 2005 sample for both sexes. The variations between age groups, however, are similar in both figures.

Comparing the results for income in figure 2.5 with those in figure 2.2, it is found that the age-specific mean net equivalent household income is higher in the 2009 sample than in the 2005 sample in all age groups and for both sexes. One may agree that this increase in incomes

[5] The 2009 microcensus is introduced in section 4.2.

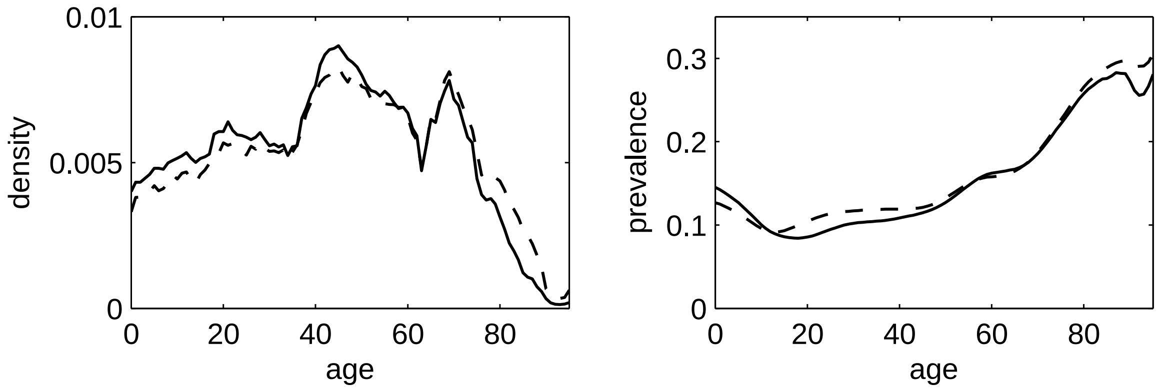


Figure 2.4: Descriptive figures (2009 sample)
Empirical density of age (left) and smoothed age-specific prevalence of sickness within the preceding four weeks (right) for males (solid lines) and females (dashed lines).

likely reflects inflation rates and increases in wages. The age-specific estimates of income inequalities in figure 2.5 do not differ noteworthy from those in figure 2.2. In summary, the results for income and income inequality are fairly similar in the 2005 and 2009 samples. As in the figures concerning density, prevalence, income and income inequality, the observed age-specific income-related health inequality in figure 2.6 is fairly similar to that in figure 2.3. One may agree that no considerable differences between both samples are evident here.

2.4.3 A note on smaller samples

As the varying inequality index introduced in this chapter has not been applied before, one may be interested in its smaller sample behavior. The following figures present estimates for the age-specific inequalities computed for random subsamples drawn from the underlying sample. The results are compared with the concentration indices computed for five year intervals as suggested by van Kippersluis et al. (2009).

Figure 2.7 presents the estimates from a 50 percent subsample ($92,780$ males and $99,611$ female). The results are similar to those in figure 2.3 and demonstrate that randomly halving the sample does not change the results considerably. The results for the 25 percent ($46,493$

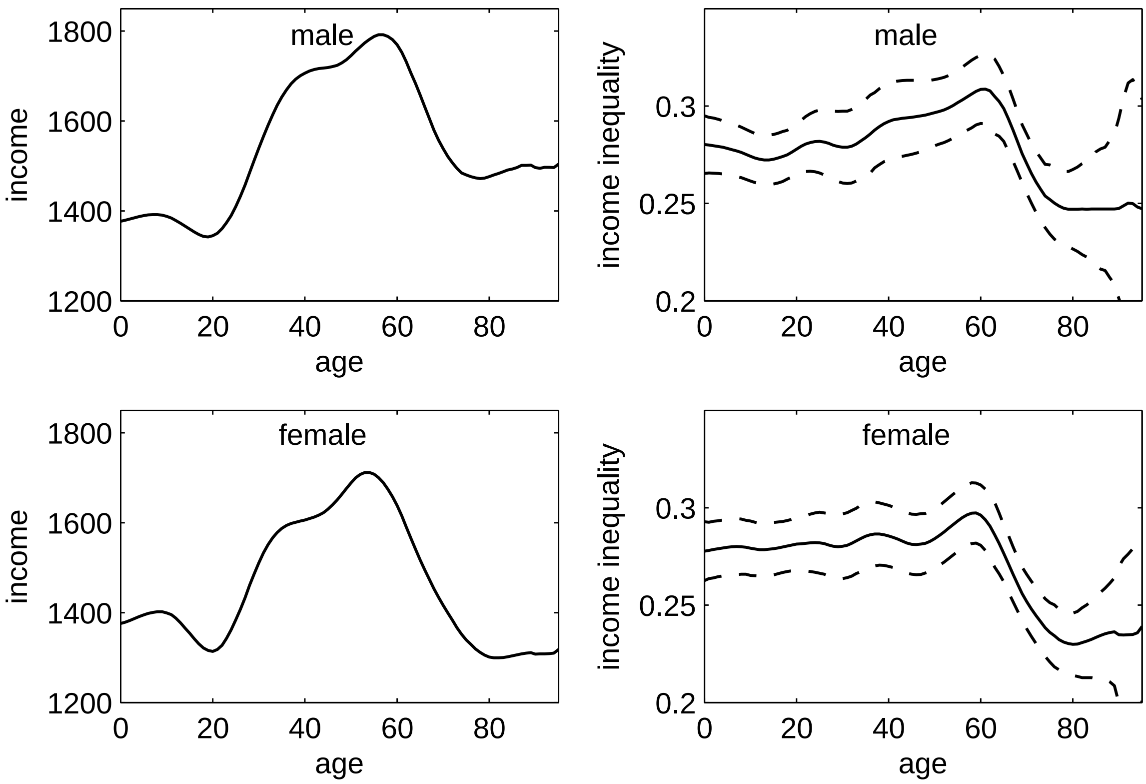


Figure 2.5: Age-specific income and Gini indices (2009 sample)
Age-specific mean (left) and Gini indices (right) with 95 percent confidence intervals (dashed lines) of net equivalent household income for males (top) and females (bottom).

males and 49,702 females) subsample in figures 2.8 also do not differ considerably from the full sample results in figure 2.3. When further reducing the sample size, the curves become flatter and variations across age groups seem less pronounced. This can be observed in the 10 percent subsample (18,523 males and 19,955 females) in figure 2.9 and the 5 percent subsample (9,252 males and 9,987 females) in figure 2.10. This flattening may to some extent be caused by a selection bias despite the random sampling from the full sample, however, the decreasing n may also have increased the degree of smoothing. It is further observed that the confidence intervals widen with decreasing sample size n.

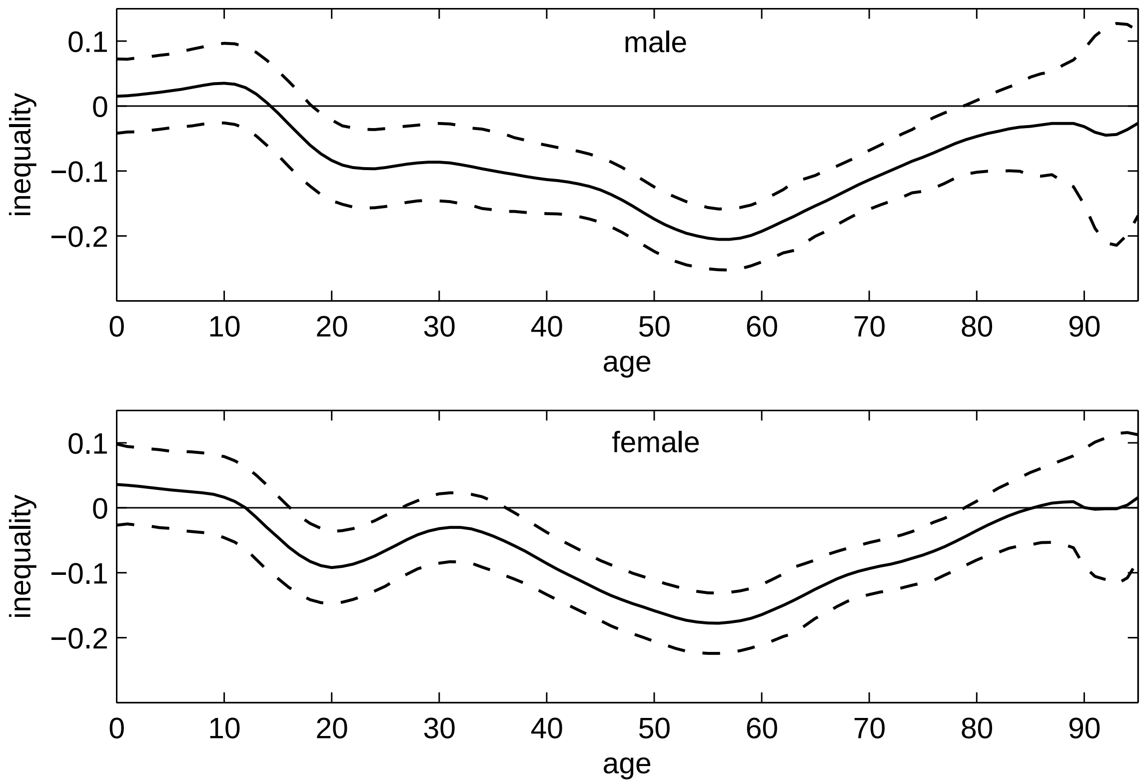
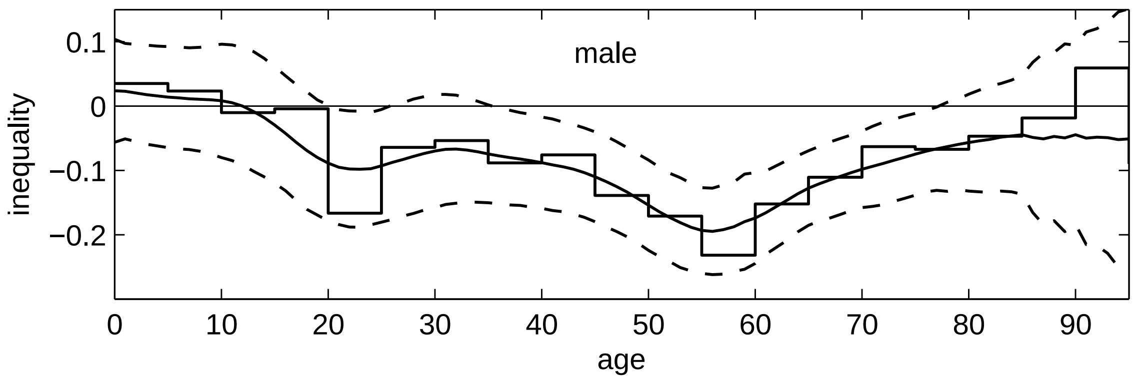


Figure 2.6: Varying inequality index (2009 sample)
Age-specific inequality (solid lines) with 95 percent confidence intervals (dashed lines) for males (top) and females (bottom).

2.4.4 A note on nonresponses

As responding to the health module was voluntary, one may be interested in the distribution of nonresponses. Figures 2.11 and 2.12 demonstrate the age-specific income-related gradient of nonresponses to the health module in the 2005 and the 2009 samples, respectively. It is found that the patterns of inequalities in nonresponses are fairly similar in both samples. However, the concentration of nonresponse among the worse-off is somewhat more pronounced and considerably more significant in the 2009 sample. Under the assumption that the economically deprived are more affected by illness than the better-off, the observed concentration of nonresponses among the worse-off may lead to some underestimation of the age-specific income-related concentration of illness among them.



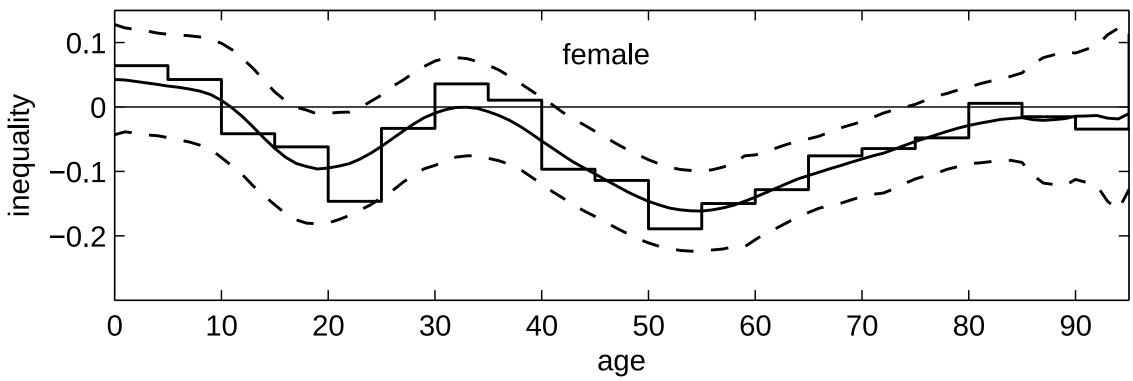


Figure 2.7: 50 percent subsample (2005 sample)
Age-specific inequality indices (smooth solid line) with 95 percent confidence intervals (dashed line) and five year interval Wagstaff indices (step function); 50 percent subsample with $92,780$ males (top) and $99,611$ female (bottom) of the 2005 microcensus.

2.5 Discussion

In this chapter, the notion of concentration indices (Erreygers, 2009; Kakwani et al., 1997; van Kippersluis et al., 2009; Wagstaff et al., 1991; Wagstaff, 2005) is combined with semiparametric regression techniques (Hastie and Tibshirani, 1993; Li et al., 2002) to a semiparametric inequality index with some convenient properties. Using the varying bandwidth inverse to local density, the index adapts itself to the data without a priori stratification into age or income groups. This method allows an age-specific computation of the inequality index with a sufficiently large number of observations guaranteed even where observations are scarce.

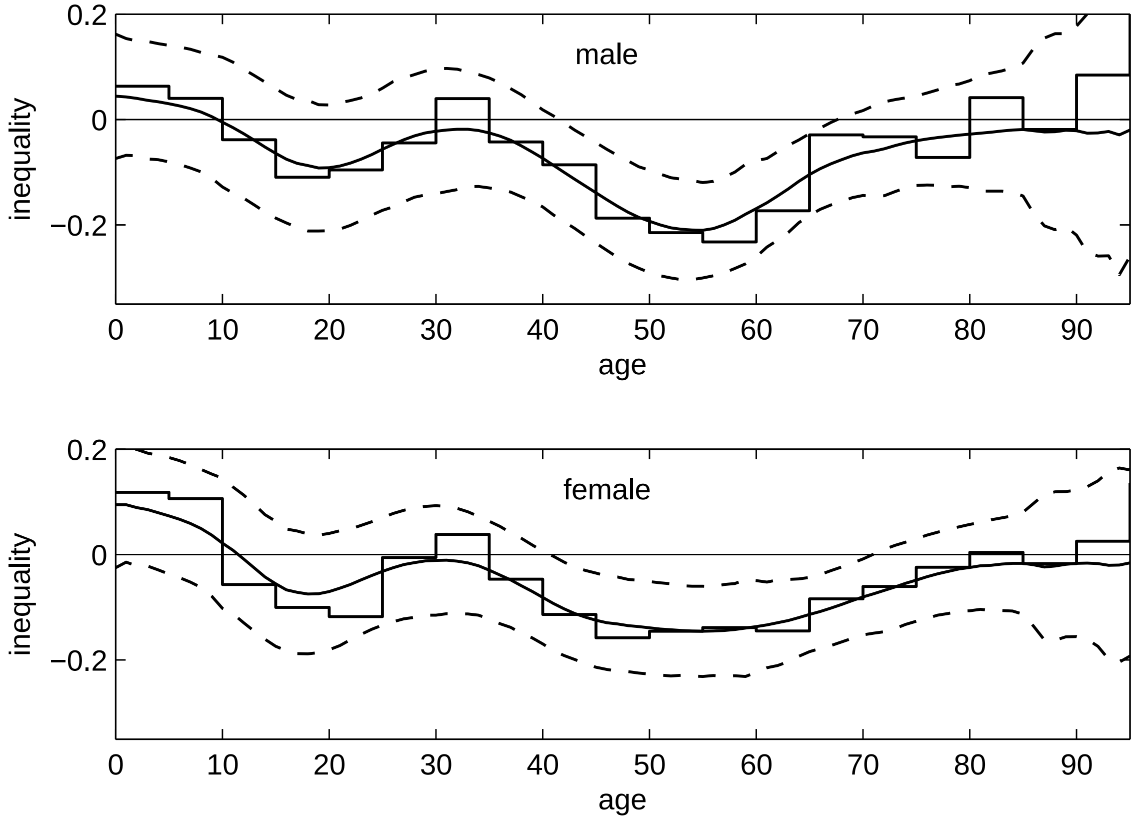


Figure 2.8: 25 percent subsample (2005 sample)
Age-specific inequality indices (smooth solid lines) with 95 percent confidence intervals (dashed lines) and five year interval Wagstaff indices (step functions); 25 percent subsample with $46,493$ males (top) and $49,702$ female (bottom) of the 2005 microcensus.

The quotient obtained through the local correction based on Wagstaff's (2005) formula allows comparisons of the extent of inequality throughout the support of the smoothing regressor. Considering the results for smaller subsamples, the overall impression is that the index performs well for samples larger than 40,000 to 50,000 observations. Where samples become considerably smaller (i.e. $< 20,000$), however, it seems that the varying inequality index should be applied with caution.

Using German microcensus data, the power of the semiparametric approach to describe age-specific income and income-related inequalities is demonstrated. Prus (2004) argues that

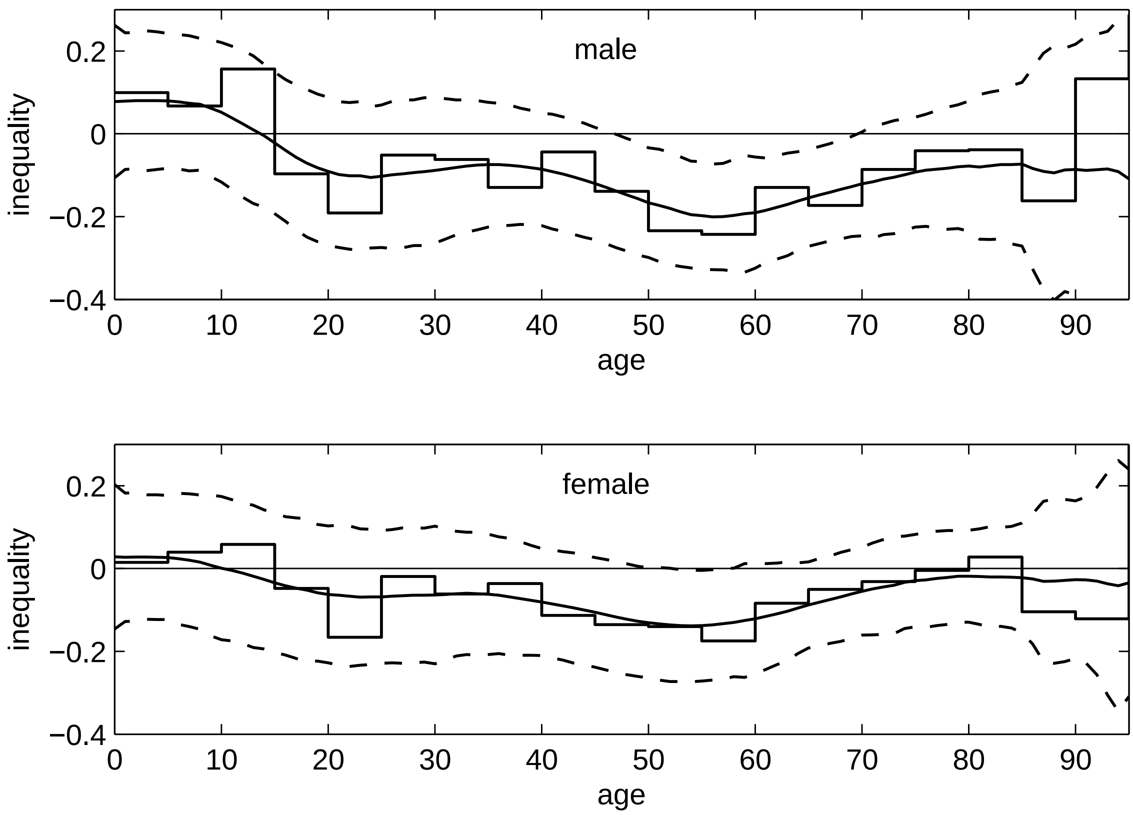


Figure 2.9: 10 percent subsample (2005 sample)
Age-specific inequality indices (smooth solid lines) with 95 percent confidence intervals (dashed lines) five year interval Wagstaff indices (step functions); 10 percent subsample with $18,523$ males (top) and $9,955$ female (bottom) of the 2005 microcensus.

one would require panel data to test the accumulation hypothesis. Similar to e.g. van Kippersluis et al. (2009), the aim of this chapter was to illustrate the variation of income and health inequalities across age groups in Germany. As a main result, it is found that direction and extent of the income-related inequality 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 70. In line with van Kippersluis et al. (2009), the strongest inequality is observed around the common age for retirement (note that the statutory age for retirement in Germany is 65, however, most people retire between 58 and 64; see Wingerter, 2010). Ac-

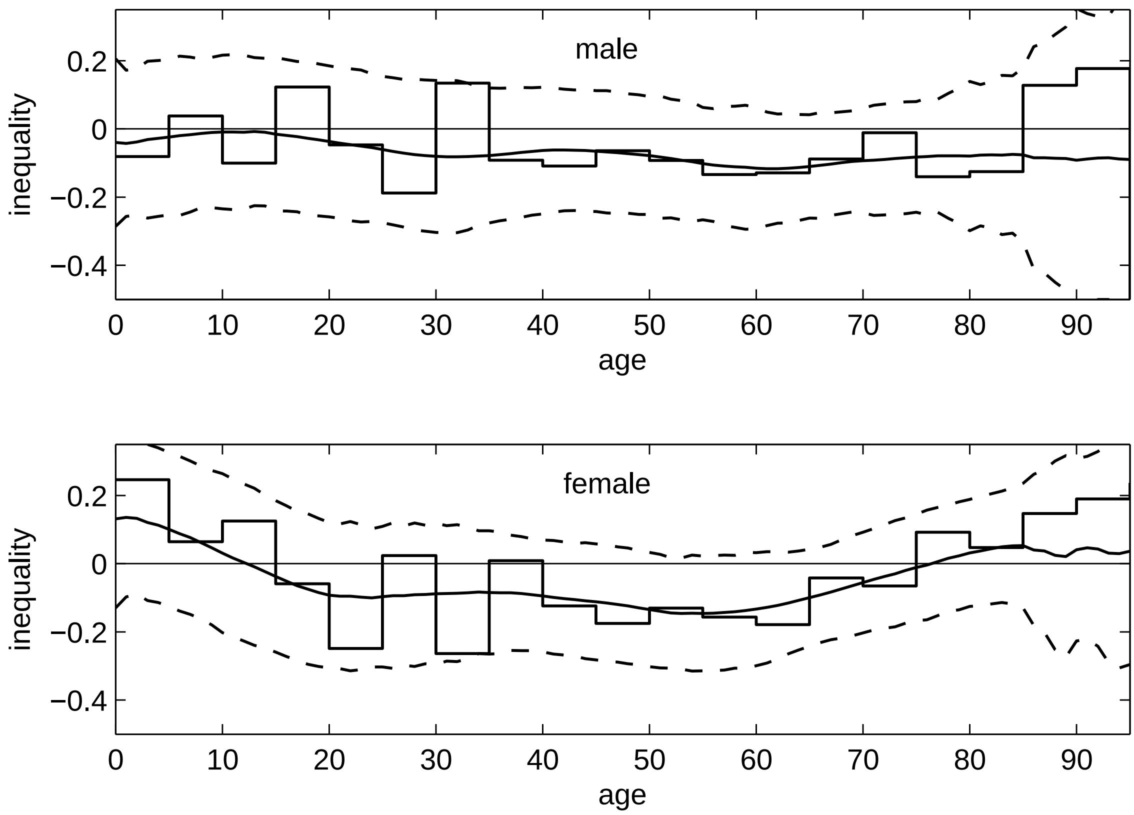


Figure 2.10: 5 percent subsample (2005 sample)
Age-specific inequality indices (smooth solid lines) with 95 percent confidence intervals (dashed lines) and five year interval Wagstaff indices (step functions); 5 percent subsample with $9,252$ males (top) and $9,987$ females (bottom) of the 2005 microcensus.

cording to Dupre (2007), the leveling found among the retired may be an artificial effect owing to mortality selection as both decline of health and increase of mortality rates are faster among the worse-off. However, Beckett (2000) has shown that this needs not necessarily be true for self-reported health.

It seems unlikely that the observed leveling is solely an artificial effect evoked by mortality selection. Mortality rates in 2005 did not exceed 2 percent before the age of 68 (74) and 5 percent before the age of 77 (81) in the male (female) sample (see Human Mortality Database, 2012).[6] The results for those older than 80 should be treated with caution, though, as mortal-

[6]Mortality rates were similar in 2009.

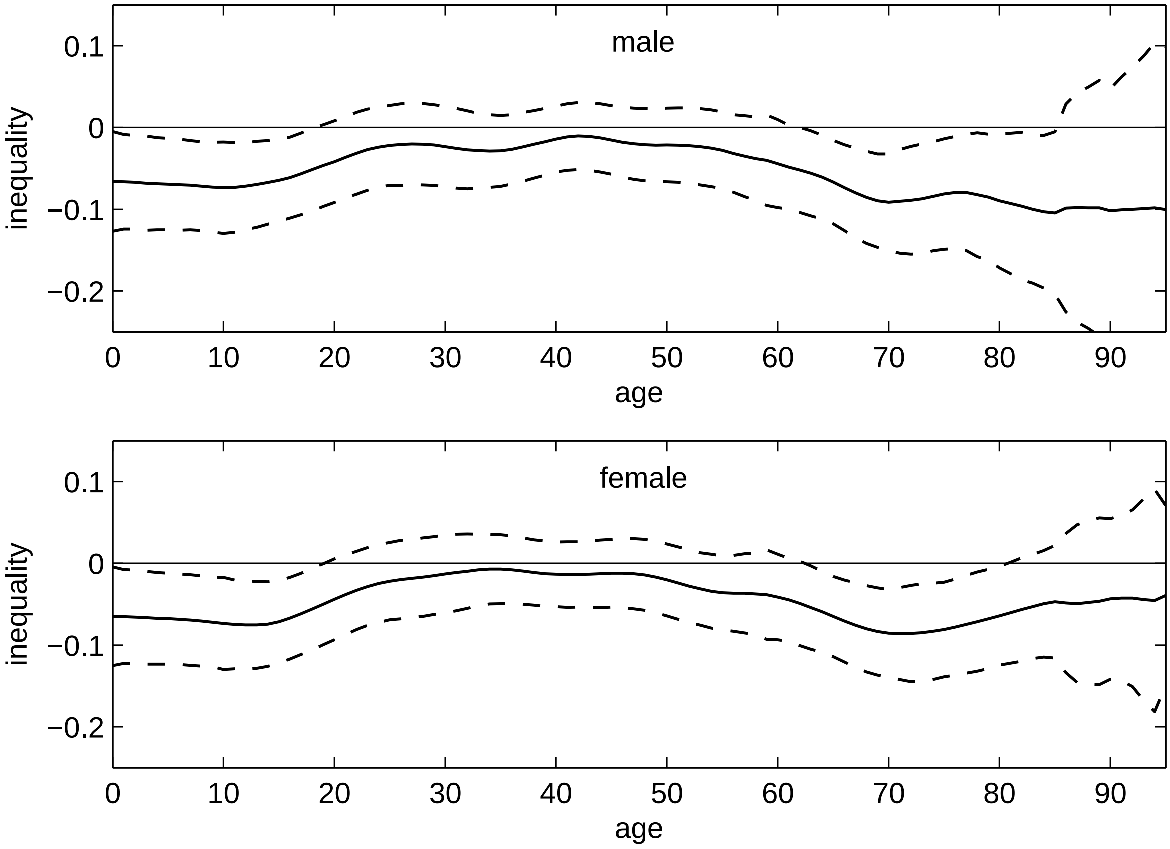


Figure 2.11: Nonresponse to health module (2005 sample)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for nonresponse in the male (top) and female (bottom) samples to the voluntary health module included in the microcensus 2005.

ity may play a considerable role in these age groups. As the age-specific income inequality persists to some extent in older age, one may further consider it as unlikely that the decline in income-related health inequalities could simply be due to an equalization of the net equivalent household income after retirement.[7]

[7]Note that technically, only income ranks, but not differences between incomes, matter. One may still consider it as possible that income equalizations may lead to a flattening of the income-related gradient if one considers income as relevant to health.

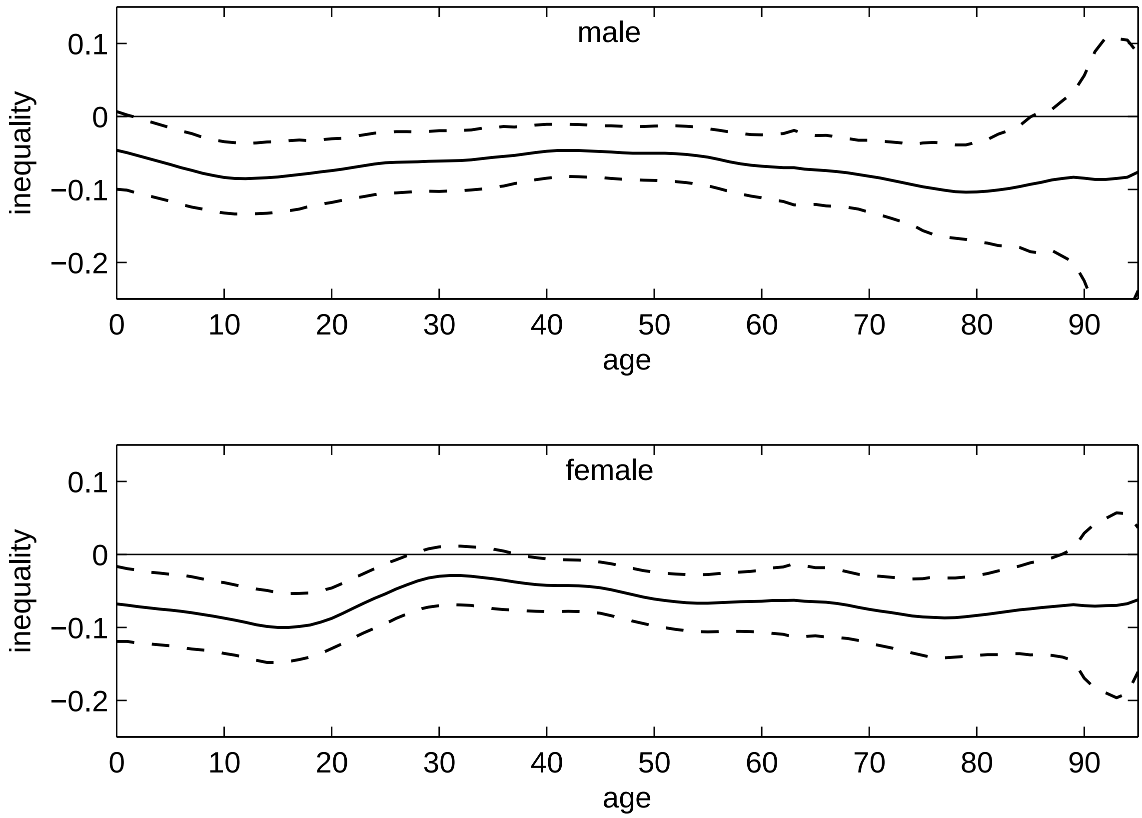


Figure 2.12: Nonresponse to health module (2009 sample)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for nonresponse in the male (top) and female (bottom) samples to the voluntary health module included in the microcensus 2005.

2.6 Conclusions

One may conclude that the observed variations in age-specific health inequalities can be seen as support the age as leveler hypothesis. The health gradient becomes pro-poor among adolescents and is strongest in the late working life. As the reduction in health inequalities coincides with a substantial increase in the age-specific prevalence, one may agree that a large part of the observed leveling can be attributed to an age-related decline of health. No income-related health inequality is observed in the female sample around age 30. As the average number of dependent children is highest around this age, one may speculate that the flatter gradient between 20 and 40 among females may be related to maternity.

Chapter 3

On age-specific variations in income-related inequalities in diabetes, hypertension and obesity[1]

3.1 Introduction

With the introduction of the health concentration index (Wagstaff et al., 1991), socioeconomic gradients in the distribution of health became an important research field in health economics (see e.g. van Doorslaer et al., 2004, 2006; Kakwani et al., 1997; McKinnon et al., 2011). Banks et al. (2007), Case and Deaton (2005) and van Kippersluis et al. (2010) point out that considering the life course perspective may be important for a better understanding of the origins of disparities in health.

The relation of low socioeconomic status and less healthy lifestyles, higher health risks and increased rates of premature mortality is well documented in the literature (Balia and Jones,

[1] A previous version of this chapter has been accepted for publication as Siegel M, Luengen M, Stock S. "On age-specific variations in income-related inequalities in diabetes, hypertension and obesity". *International Journal of Public Health*, **DOI:** 10.1007/s00038-012-0368-7; downloadable from `http://www.springerlink.com/content/181l52078524547u/`.

2008; Case and Deaton, 2005; Deaton and Paxson, 1998; Kim and Durden, 2007; Lynch, 2003; Prus, 2004). The two currently predominant explanations for variations in socioeconomic health gradients over the life course are the disadvantage accumulation hypothesis and the age as leveler hypothesis. The former states that social health gradients develop in early life and emerge over the complete life course through continual accumulation of health disadvantages among the deprived (Kim and Durden, 2007; van Kippersluis et al., 2010; Lynch, 2003; Prus, 2004; Ross and Wu, 1996). The age as leveler hypothesis, on the other hand, has long been considered as an antithesis (Dupre, 2007). It is, however, rather an extension adding the assumption that the decline of health (being an inevitable part of aging) may outperform the accumulation of health disadvantages around mid-age (Case and Deaton, 2005; Deaton and Paxson, 1998; Kim and Miech, 2009; Kunst and Mackenbach, 1994). According to van Kippersluis et al. (2010), evidence for the age as leveler hypothesis is generally consistent with an accumulation process up to some point where age begins acting as a leveler in later life. Associating the accumulation process to the individual and leveling through age to the aggregate level, Dupre (2007) has shown that both mechanisms are by no means contradictory.

A growing body of literature analyzes the health status over the life course for distinct socioeconomic groups. van Kippersluis et al. (2010), for instance, use cross sectional data from the Netherlands and compare the development of self reported health, disability and mortality for distinct educational groups. Mirowsky and Ross (2005) investigate differences in the predicted health status between educational levels in the United States. Both find that health status declines faster among the less educated. Using data from eleven European countries, van Kippersluis et al. (2009) find a continual disadvantage accumulation for most countries, however, they discover some leveling for France, Germany and the U.K.

No work has yet been done on the life course perspective in income-related inequalities of specific diseases. The aim of this chapter is to highlight changes of the social gradients in three health outcomes across different age groups. Using German survey data, age-specific variations in income-related health inequalities in hypertension and diabetes are investigated. The risk of hypertension and diabetes is negatively correlated with socioeconomic status in developed countries and both are considered as avoidable through lifestyle and health behavior (Harati et al., 2010; Puska, 2010). The present chapter further analyzes obesity as it increases the risk of hypertension and diabetes considerably (Haffner, 2006; WHO, 2003). The semi-parametric extension of the concentration index introduced in chapter 2 is used to measure age-specific income-related health inequalities.

3.2 Data

To assure sufficiently large numbers of observations particularly in the oldest age groups, data from the 2002 and 2006 waves of the Health Care Access Panel are pooled for the empirical analysis. The data were collected by TNS Healthcare, Munich (now Kantar Health) for commercial purposes. For each wave of this voluntary mail survey, approximately $50,000$ households were drawn randomly from a large database. In-field time was 6 weeks without active reminding and the response rates were approximately 60 percent in both waves. Potthoff et al. (2004) provide an extensive discussion on the 2002 wave of the HCAP showing that the underlying sampling procedure yields representative samples for the German population.

The pooled sample comprises $117,167$ individuals in $48,574$ households. $75,122$ individuals in $29,421$ households were included in 2002 and $60,555$ individuals in $28,828$ households in 2006. $18,510$ individuals in $8,718$ households participated in both surveys. As chronic conditions rarely affect younger individuals and a meaningful interpretation of the body mass index is problematic for children and adolescents, $28,390$ individuals younger than 20 were removed. Another $1,176$ observations were dropped owing to missing data on income. The sample for the empirical analysis eventually comprises $87,601$ individuals ($45,889$ women and $41,712$ men) in $47,867$ households.

The social status is measured using the monthly net equivalent household income. In the

HCAP, individuals were asked to specify the sum of all incomes available to the household (e.g. from labor, capital, pensions or welfare benefits). The modified OECD equivalence scale is used to compute net equivalent household income (see e.g. van Doorslaer et al., 2004; van Kippersluis et al., 2009). Benefits from the public pension scheme depend largely on incomes over the life course (Boersch-Supan and Wilke, 2004) and German welfare policies are considered as highly status preserving (Brockmann et al., 2009). Disparities in the income distribution and the relative socioeconomic position are therefore likely to persist after retirement. Accordingly, one may consider household income as a good measure of the social position before and after retirement.

The body mass index (BMI) is computed from self-reported anthropometric data as body weight in kg divided by the squared body height in meters, kg/m^2. Following the standard WHO classification (WHO, 2009), obesity is defined as BMI over 30. The second health outcome is hypertension. Individuals were asked to indicate whether they had hypertension during the preceding twelve months. The third health outcome is diabetes in the preceding twelve months, however, the survey does not allow a distinction between type one and type two diabetes. Type two diabetes is agreed to be age related and influenceable through lifestyles and health behavior (see e.g. Harati et al., 2010; Puska, 2010). Diabetes is analyzed irrespective of insulin-dependency because type one diabetes is here considered as mainly genetic and therefore equally distributed across socioeconomic groups.

3.3 Econometric Model

Age-specific income-related health inequalities are measured using the semiparametric extension of Wagstaff's (2005) corrected concentration index. Concentration indices are derived from concentration curves where the cumulative share of health is compared with the cumulative share of the population ranked by income (Wagstaff et al., 1991). If no income-related inequality is present, the concentration curve coincides with the line of equality (45° line).

Technically, the concentration index measures twice the area between the concentration curve and the line of equality. It is bound in the $(-1;1)$ interval and becomes positive (negative), if the variable of interest concentrates among the better-off (worse-off). The index equals zero if no income-related inequality is observed. Note that the concentration index is a bivariate extension of the Gini index; using income in place of health as outcome variable would yield the well known Lorenz curve and Gini index. Konings et al. (2010) provide an intuitive introduction to the concept of Gini-type concentration indices.

The model proposed in chapter 2 is used to obtain the concentration index as a smooth function of some regressor z (i.e. age). The convenient regression approach (Kakwani et al., 1997) is adapted to estimate

$$2\frac{\sigma_r^2(z)}{\mu(z)}y = \beta_0(z) + \beta_1(z)r(z) + \varepsilon, \tag{3.1}$$

using the varying smooth coefficient approach (Hastie and Tibshirani, 1993; Li et al., 2002). Individuals $i = 1,...,n$ must be sorted by income in ascending order and $r(z)$ denotes the local fractional rank. The coefficient $\beta_1(z)$ is the varying concentration index $C(z)$ and y is the dependent variable with local mean $\mu(z)$.

As proper sample weights are not available for the pooled data, formula (2.5) from section 2.2.3 for the fractional rank simplifies to

$$r_i(z) = \sum_{j=1}^{i} k_{h_z}\left(u_j\right) - \frac{k_{h_z}(u_i)}{2} \tag{3.2}$$

with $u_i = z_i - z$. Using the kernel weights $k_{h_z}(u_i)$ is important here to assure that the local estimates for the mean and variance of the local rank variable $r(z)$ equal their theoretical asymptotic values 0.5 and $1/12$ throughout the support of z. A Nadaraya-Watson estimator with a quartic kernel is used. This kernel function assigns higher weights to observations closer to z, lower weights for observations being further away and zero weight to observations outside the bandwidth h_z. The kernel weights are $k_{h_z}(u_i) = \left(\sum_{j=1}^{n} K_{h_z}(u_j)\right)^{-1} K_{h_z}(u_i)$ with $K_{h_z}(u_i) = (1/h_z)\,K(u_i/h_z)$ and $\sum_{i=1}^{n} k_{h_z}(u_i) = 1$. The local bandwidth parameter h_z is chosen inversely to the local kernel density f_z owing to the tradeoff between estimation bias and uncertainty. The kernel density is the nonparametric estimate of the probability density function (pdf) of z.

Comparing concentration indices of binary variables with varying means may be misleading as the bounds then depend inversely on the mean μ, $|C|_{max} = 1-\mu$ (Wagstaff, 2005, 2011). Alike in chapter 2, Wagstaff's (2005) approach is adapted such that $W(z) = C(z)/\left(1-\mu(z)\right)$ is the varying Wagstaff index. Similar to its homogeneous counterpart, $W(z)$ is always bound in the $(-1;1)$ interval irrespective of $\mu(z)$. This allows comparisons of inequalities throughout the support of z. The varying Wagstaff indices are computed separately for men and women and pointwise confidence intervals are reported.

The local standard error $\sigma_C(z)$ of the varying concentration index can be approximated by estimating β_0^* and β_1^* from equation (3.1) with the untransformed health variable y in place of $\left(2\sigma_r^2(z)/\mu(z)\right)y$. One may then take advantage of the fact that $\beta_0^*(z)+0.5\beta_1^*(z) = \mu(z)$ and apply the δ method to $C(z) = 2\sigma_r^2(z)\beta_1^*(z)\left[\beta_0^*(z)+0.5\beta_1^*(z)\right]^{-1}$ (Kakwani et al., 1997; Wildman, 2003). The standard error $\sigma_W(z)$ for $W(z)$ can be estimated analogously (see chapter 2). Note that mean and variance of the rank variable need not be considered as stochastic as they are sample independent. The error term may likely be heteroscedastic and autocorrelated (Kakwani et al., 1997; Wildman, 2003; McKinnon et al., 2011); local Newey-West type standard errors are therefore estimated (for computational details and a more technical introduction of the method, see chapter 2).

3.4 Results

The graphs for the estimated density $\hat{f}(z)$ with respect to age (the smoothing parameter z) in figure 3.1 exhibit similar patterns for both sexes. Cohorts younger than 35 are smaller than

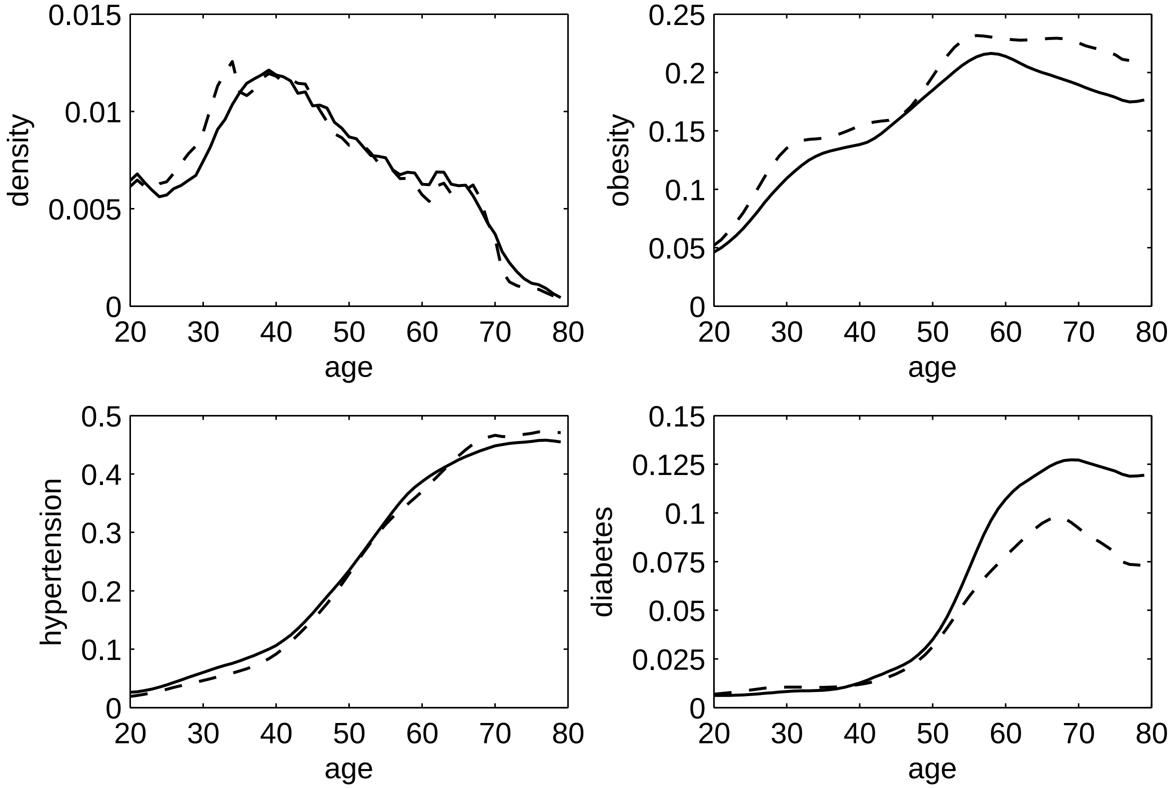


Figure 3.1: Kernel density and age-specific prevalences (pooled sample 2002, 2006)
Empirical density of age (upper left) and age-specific prevalences of obesity (upper right), hypertension (bottom left) and diabetes (bottom right) for the male (solid lines) and female (dashed lines) samples from the 2002 and 2006 Health Care Access Panel.

those born before the 1970's. This corresponds with the population pyramid for Germany and supports existing evidence for an aging society (see e.g. von Weizsäcker, 1996). Note that the observed kernel density $\hat{f}(z)$ implies that the bandwidth parameter h_z is minimal at age 33 for women ($h_z \approx 5.95$) and at 39 for men ($h_z \approx 6.22$). The bandwidth is largest for the 79 years old for both sexes ($h_z \approx 16.6$ for women and ≈ 16.9 for men).

The overall sample prevalence of obesity in table 3.1 is higher among women than among men. Comparing the age-specific estimates in figure 3.1, it is found that this holds in all age groups. The prevalence is highest among the 56 years old women and the 58 years old men and lower in the younger and older age groups. The overall-sample prevalence of hypertension

Table 3.1: Descriptive statistics (pooled sample 2002, 2006)

	female ($n = 45,889$)			male ($n = 41,712$)		
	prevalence	$\hat{W}$	$\hat{\sigma}_W$	prevalence	$\hat{W}$	$\hat{\sigma}_W$
income	$1,394.90^a$	0.2971^{*b}	0.0032^c	$1,460.90^a$	0.2994^{*b}	0.0035^c
obesity	16.57%	-0.1283^*	0.0126	15.19%	-0.0543^*	0.0137
hypertension	17.73%	-0.0623^*	0.0122	19.54%	-0.0114	0.0123
diabetes	3.14%	-0.1283^*	0.0273	4.15%	$-0.0535^\dagger$	0.0246

Homogeneous estimates from the 2002 and 2006 Health Care Access Panel.
*) significant at the 1 percent level; †) significant at the 5 percent level;
^{a}) mean of net equivalent household income; ^{b}) Gini-index of net equivalent household income (without Wagstaff's correction); ^{c}) standard error of the Gini-index

is higher among men. The shape of its age-specific prevalence curve in figure 3.1 suggests that prevalence rises almost monotonously with increasing age and flattens after age 70 for both sexes. It is highest among the 76 years old women and 77 years old men. The overall sample prevalence of diabetes in table 3.1 is again somewhat higher for men. The age-specific estimates vary around one percent being somewhat higher for women until the late thirties. The prevalence then rises faster among men, is highest at age 69 ($\approx$ 12 percent) and varies around this value for the oldest. For women, it is highest at age 67 ($\approx$ 10 percent) and decreases to approximately 7 percent for the oldest.

The age-specific mean net equivalent household income in figure 3.2 yields similar patterns for men and women. The graphs suggest that the youngest individuals live in households with the lowest incomes. A first peak is observed around the age of 30 and the highest mean income is found shortly before the statutory retirement age of 65. The lower mean net equivalent household incomes between 35 and 50 stem from an increase in the equivalence weights owing to a peak in the average number of children per household in these age groups. In all age groups, men have a higher mean income than women. The age-specific Gini index is somewhat lower for women and men around 30 than for the youngest. After age 30, income

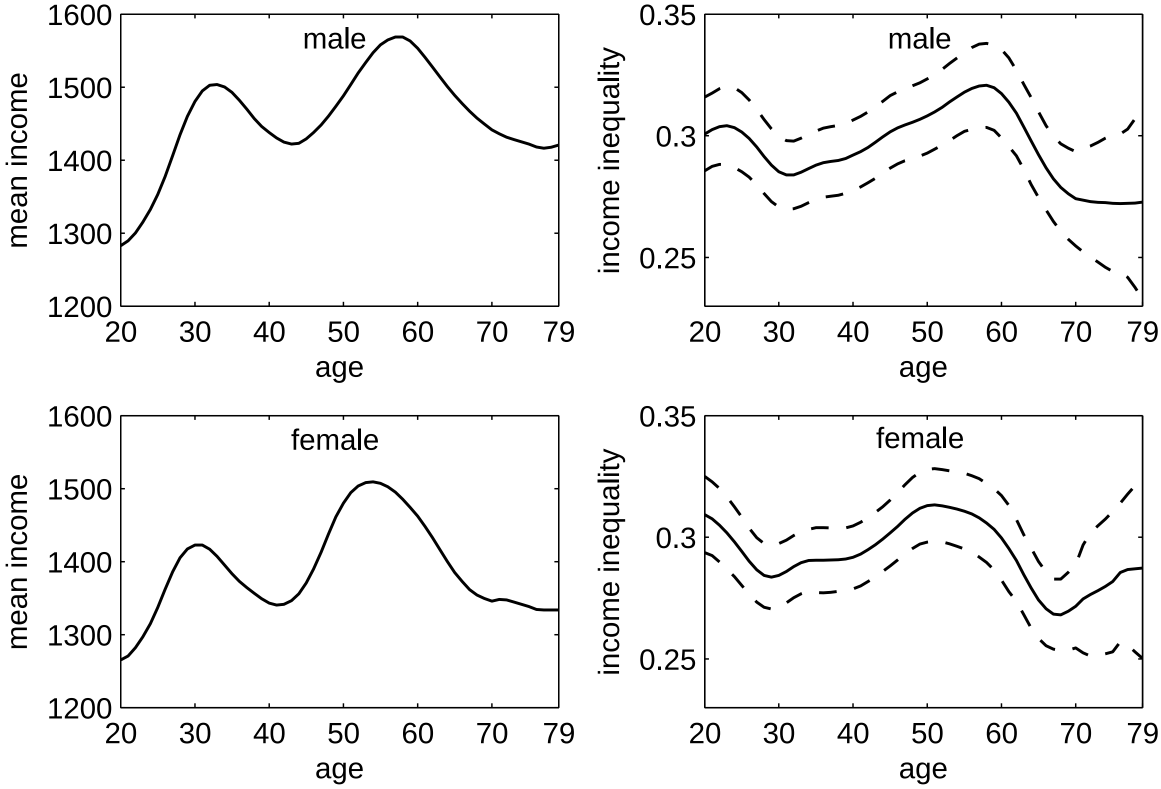


Figure 3.2: Age-specific income and Gini indices (pooled sample 2002, 2006)
Age-specific mean net equivalent household incomes (left) and Gini indices (right) with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel.

inequality increases and is highest for women around 50 and men around 60. The Gini index drops back to approximately 0.27 for both sexes during the common retirement period. Note that the statutory retirement age is 65, however, most people retire in the preceding seven years (see Wingerter, 2010). Although the results suggest that retirement leads to some leveling in household incomes, considerable income inequalities persist after retirement.

The homogeneous Wagstaff indices W for obesity in table 3.1 are both significantly negative suggesting a concentration of the disease among the worse-off for both sexes. The varying Wagstaff index for men in figure 3.3 is negative but statistically significant for no age. The significantly negative index for women aged between 28 and 72 indicates a concentration among

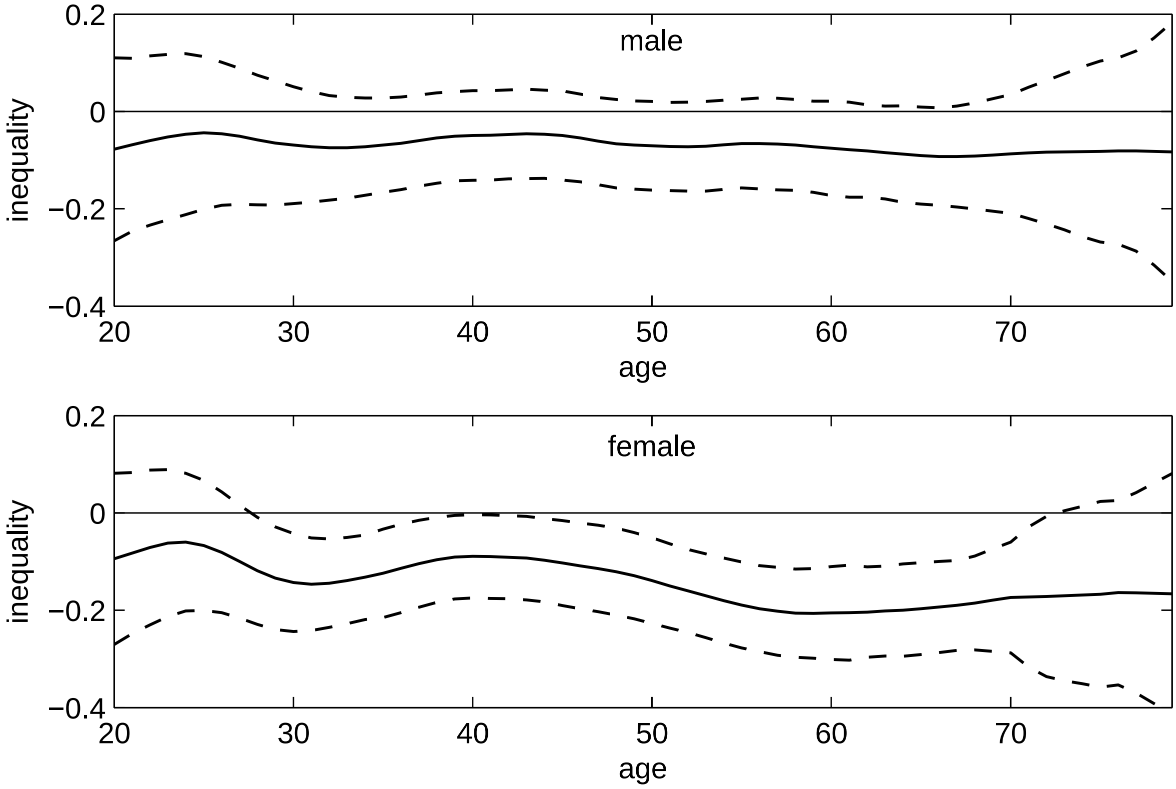


Figure 3.3: Age-specific inequality of obesity (pooled sample 2002, 2006)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel.

the worse-off in these age groups. While the varying index for men shows no considerable changes over the life course, women exhibit a lower level of inequality in mid-age and a higher level among those over 55.

The homogeneous Wagstaff index for hypertension among women in table 3.1 is again highly significant. The varying Wagstaff index in figure 3.4 is statistically significant for those between 50 and 69 in the female sample. For all age groups, the index varies around -0.1. The varying index for men is negative for those aged 23 or older. The homogeneous and varying Wagstaff indices for the male sample are both insignificant.

The homogeneous Wagstaff indices for diabetes in table 3.1 are significantly negative for

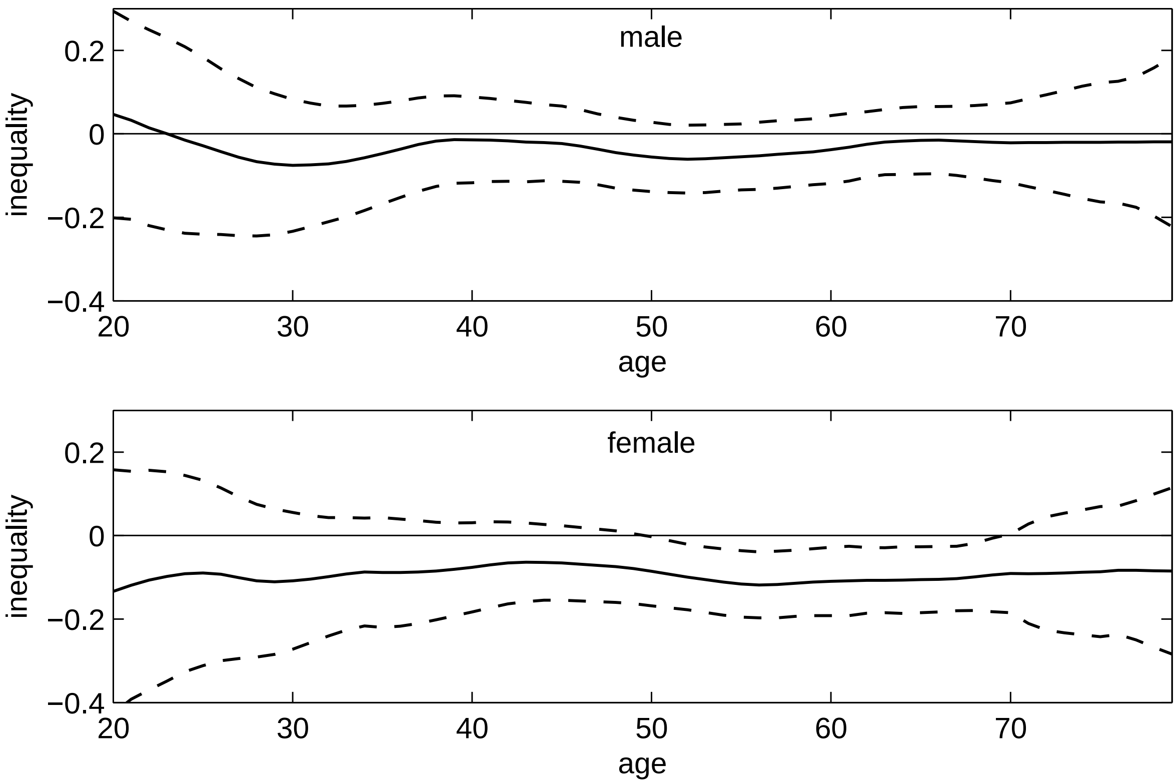


Figure 3.4: Age-specific inequality of hypertension (pooled sample 2002, 2006)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel.

women and men, again suggesting a concentration of the disease among the worse-off. The varying Wagstaff index for the female sample in figure 3.5 is negative for all age groups. It is, however, only significant for women aged between 51 and 62. No significant age-specific income-related gradients are observed for diabetes in the male sample.

3.5 Discussion

Using the varying inequality index introduced in chapter 2, this chapter addressed age and sex specific variations in income-related health inequalities. One may consider computing batteries of concentration indices for distinct age groups (as e.g. in van Kippersluis et al.,

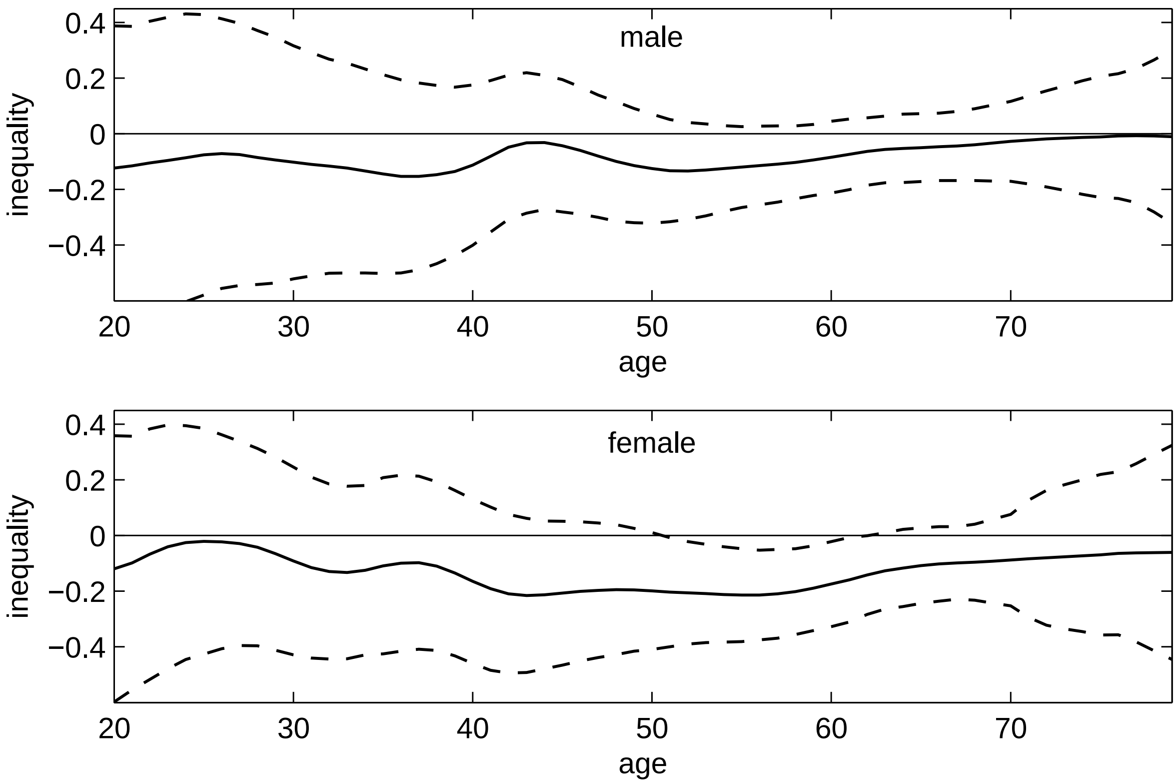


Figure 3.5: Age-specific inequality of diabetes (pooled sample 2002, 2006)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel.

2009) and smoothing over these as an alternative. As argued in chapter 2, however, this would involve double uncertainty and a considerable bias. First, particularly the oldest age groups may become rather small which would involve high uncertainty and large confidence intervals. Second, smoothing over such estimates would add its own uncertainty. In addition, the estimates close to the age group limits would likely be biased which is avoided in the kernel smoothing approach.

This chapter has shown that socioeconomic gradients in obesity, hypertension and diabetes vary across age groups in Germany. Significant age-specific inequalities to the detriment of the deprived in hypertension and diabetes are only observed for women in midlife. The age-specific gradient for obesity is significant for almost all age groups in the female sample.

Somewhat surprisingly, no significant age-specific income-related inequalities were found among men. Although some leveling is observed for diabetes among both sexes and hypertension among men, one may agree that the results are neither a clear support for the accumulation nor for the age as leveler hypothesis.

Using the varying index has some advantages compared with computing homogeneous concentration or Wagstaff indices. Using the latter may invoke what Islam et al. (2010) refer to as student and pension effects. According to the former, young people in the beginning of their career, in job trainings or universities are usually in good health conditions but have low incomes. This may result in an underestimation of health disadvantages for lower income households. Conversely, the pension effect describes the notion that older people have comparably bad health conditions due to their age and lower incomes after retirement. This, again, may lead to an overestimation of health disadvantages for the poor. The homogeneous indices actually underestimate the extent of inequality for most age groups in hypertension and obesity compared with the age-specific results.

When considering self-reported data on specific diseases, one may argue that certain diseases have to be diagnosed by a physician. Individual awareness may therefore, to some extent, depend on health care utilization and communication between patients and physicians. However, approximately 90 percent of the German population contact a physician within a year and, more importantly, Germany is known for a fairly equitable access to health care (see e.g. van Doorslaer et al., 2004, 2006). Considerable reporting biases owing to inequalities in health care utilization are hence rather unlikely. The potential of biases owing to social distances between physicians and less educated or lower income patients, however, remains. Considering the results found in Kelly-Irving et al. (2011), one may speculate that such biases may lead to a pro-rich bias, i.e. an underestimation of inequalities to the detriment of the deprived. Concerning the results for obesity, it should be mentioned that self-reported anthropometric data may involve some measurement or reporting bias which may lead to an underestimation of the prevalence and socioeconomic gradient of obesity.

Using cross sectional data involves some limitations. First, they do not facilitate a clear distinction between life course effects and cohort effects. However, van Kippersluis et al. (2009) found no consistent cohort effects for Germany. Although cohort effects cannot fully be excluded, one may agree that these likely play a minor role and the observed variations are mostly an age effect. Second, the data do not allow a distinction between income effects on health and health effects on income or to test the causal relationship between then. However, one may assume possible losses of income through adverse health related selection into both early old age and reduced earning capacity retirement to be of minor importance (Brockmann et al., 2009). Only 1.5 and 1.6 percent of women and men among the 45 to 50 years old in the sample are retired. The shares are somewhat higher among the 50 to 55 years old; 2.8 and 4.3 percent, respectively. Further, German welfare policies are seen as rather status preserving (Brockmann et al., 2009). Pensions largely depend on compulsory contributions over the life course (Boersch-Supan and Wilke, 2004) The relative socioeconomic position within a birth cohort will therefore change only marginally with retirement. Finally, Dupre (2007) and Prus (2004) argue that leveling may be an artificial result owing to mortality selection. However, Beckett (2000) has shown that this needs not necessarily be true for self reported health. The age-specific mortality rates did not exceed 0.5 (1) percent for women younger than 60 (65) in the respective years in Germany (Human Mortality Database, 2012). The rates were about double for men. One may therefore reject the notion that the observed variations in income-related inequalities could be solely caused by mortality selection.

3.6 Conclusions

One may agree that the results are of particular interest for researchers and health policy makers alike since the health outcomes under consideration are among the major risk factors for cardiovascular diseases and premature mortality in developed countries. In addition, this chapter adds an important contribution to the field of age-specific analyses of income-related health inequalities. Given the demographic developments in most industrialized countries as demonstrated by the evidence for an aging German society found in the data, such analyses will likely become an increasingly important tool for health policy makers to maintain an efficient allocation of both, preventive and curative health services.

Chapter 4

Age-specific variations in income-related inequalities in smoking behavior

4.1 Introduction

A broad body of literature provides evidence on the various hazardous effects of smoking. For example, smoking increases the risk of pulmonary and cardiovascular diseases (Kamholz, 2004), asthma among adolescents (Genuneit et al., 2006) and lung cancer (Peto et al., 2000). It is associated with premature mortality (Balia and Jones, 2008) and lower quality of life (Slama, 2008).

Social gradients in smoking behavior in industrialized countries are well documented in the literature. Individuals with lower socioeconomic status, in general, have higher consumption levels, start smoking earlier in life and are less likely to quit (Schaap and Kunst, 2009). These finding were repeatedly confirmed for Germany. Lampert and Burger (2004) compare the prevalence of smoking in different social classes using the 2003 German National Telephone Health Survey. Helmert and Buitkamp (2004) analyze survey data from four national health surveys and three waves of the Bertelsmann Health Monitor. Both find that tobacco consumption is less common among people with higher education, occupational status and income. Analyzing data from the 1998 German National Telephone Health Survey, Lampert and Thamm (2004) find that both educational and income gradients in smoking prevalence reduce with increasing age.

Giskes et al. (2005), Graham (1996) and Schulze and Mons (2006) describe the development of the smoking epidemic since the early twentieth century and find similar patterns in most industrialized countries. Smoking was rather uncommon among women and prevalence rates first rose among higher educated men. While becoming more common among the less educated, smoking prevalence declined among individuals with higher social status. According to Schulze and Mons (2006), the social gradient reversed with the 1930s birth cohort. Lampert (2010) addresses the recent developments in Germany in the light of actions undertaken to reduce the smoking prevalence. Alike Giskes et al. (2005), however, he finds no significant changes over the recent years. Using data from twelve European countries, Graham (1996) find leveling in gender differences in smoking behavior due to a faster decline of the smoking prevalence among men.

To quit smoking may prevent the incidence of smoking-related diseases even in later midlife (Peto et al., 2000). Monitoring socioeconomic inequalities in tobacco consumption is therefore crucial when aiming at an equitable distribution of health (Schaap and Kunst, 2009). Empirical evidence concerning age-specific variations in income-related inequalities in smoking behavior, however, still seems scarce. This chapter measures age-specific income-related inequalities in smoking behavior based on three outcome variables drawn from the German microcensus. The outcomes are current smoking, ever (former or current) smoking and smoking cessation among the ever-smokers. As analyzing social gradients in smoking behavior based on homogeneous measures would neglect the above described developments even when controlling for birth cohorts, the semiparametric extension of the concentration index introduced in chapter 2 is applied.

4.2 Data and variables

Data are drawn from the 2009 wave of the German microcensus which is a representative survey of the German population conducted by the Research Data Centers of the Federal Statistical Office and the Statistical Offices of the Federal States (Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder). Comprising approximately one percent of the German households, the microcensus is considered as the most representative survey available for Germany (FSO, 2006; Reeske et al., 2009). Households are included in four consecutive surveys and 25 percent of the households are replaced each year.

The present chapter uses the scientific use file comprising a random subsample of approximately 70 percent ($n = 489,349$) of the German microcensus. Children younger than 15 ($64,808$) were not asked about smoking and $80,968$ individuals over 15 did not respond to the smoking-related questions. $343,573$ individuals ($179,659$ female and $163,914$ male) aged 15 or older provided information about their smoking behavior. Another $26,597$ observations ($12,763$ male and $13,834$ male) had to be removed because of missing information on household income. The final sample comprises $316,976$ ($151,151$ male and $165,825$ female) individuals. Sample weights were adjusted accordingly.

Interviewees were asked whether they currently were frequent, occasional or non-smokers. Non-smokers were then asked whether they were former frequent or occasional smokers. As one may agree that the subjective distinction between frequent and occasional smoking is rather weak, individuals are only grouped into smokers and non-smokers and smoking is analyzed regardless of its frequency. The first outcome variable is current smoking, the second outcome is ever-smoking. Ever-smokers are individuals who either currently smoke or formerly smoked. As a third outcome, smoking cessation is measured as former smoking among the ever-smokers. It should be noted here that this measure of smoking cessation does not only include those who stopped smoking in the respective year and may thus overestimate the annual cessation rates.

The socioeconomic status is measured using household income. Data on income comprise all possible sources (e.g. from labor, capital or pensions). To account for the household size, total income is adjusted using the modified OECD equivalence scale (see e.g. van Doorslaer et al., 2004). One may question the reliability of income as a measure for the socioeconomic position particularly after retirement. While this may be problematic for some countries, the German welfare system is considered to be highly status preserving (Brockmann et al., 2009). Approximately 90 percent of the German population are covered by the public pension scheme and benefits depend largely on compulsory contributions until retirement (Boersch-Supan and Wilke, 2004). One may therefore agree that the relative socioeconomic position within one's age group is unlikely to change considerably with retirement. Income is thus used to measure the socioeconomic status in all age groups.

4.3 Methods

Age-specific income-related inequalities are measured using the semiparametric extension of Wagstaff's corrected concentration index (Wagstaff, 2005, 2011) introduced in chapter 2. Let the outcome be y with mean μ. The weighted fractional rank is $r_i = \sum_{j=1}^{i} w_j - {}^{w_i}/_2$ with $\sum_{i=1}^{n} w_i = 1$ and individuals $i = 1, \ldots, n$ sorted in ascending order by income. The concentration index is $C = {}^2/_\mu \sum_{i=1}^{n} y_i r_i - 1$ and measures twice the area between the line of equality and the concentration curve (Wagstaff et al., 1991). The latter plots the cumulative share of outcome y against the cumulative share of the population ranked by income. C is bound in the $(-1;1)$ interval and becomes positive (negative), if the outcome is concentrated among the rich (poor). Where no inequality is observed, the concentration curve coincides with the line of equality and C equals zero (see e.g. van Doorslaer and Koolman, 2004; van Doorslaer et al., 1997, 2004, 2006; Wagstaff et al., 1991; Wagstaff, 2005, for descriptions, discussions and applications). Konings et al. (2010) and O'Donnell et al. (2008) provide intuitive introductions to the concept of Gini-type concentration indices.

In chapter 2, the convenient regression approach (Kakwani et al., 1997) has been combined with the varying coefficient model (Li et al., 2002) to obtain the concentration index as a smooth function of some regressor $z \in Z \subset \mathbb{R}$. The convenient regression formula

$$2\frac{\sigma_r^2(z)}{\mu(z)}y = \beta_0(z) + \beta_1(z)r(z) + \varepsilon \tag{4.1}$$

and the z-specific mean $\mu(z)$ of y can be estimated using nonparametric estimation techniques. The z-specific concentration index is then $\beta_1(z) = C(z)$, $\sigma_r^2(z)$ is the variance of $r(z)$ and ε denotes the error term. Note that if y is the living standard variable used for ranking, C is a varying Gini index.

As pointed out in chapter 2, the sample and kernel weights w and $k_{h_z}(u)$ must be included in the computation of the fractional rank. Technically, the condition that the locally weighted mean and variance of $r(z)$ must be 0.5 and $1/12$, respectively, must hold for any $z \in Z$. Intuitively, when computing the concentration index for a specific z, including all observations regardless of z would be misleading. Alike in equation (2.5), $r(z)$ is computed as

$$r_i(z) = \sum_{j=1}^{i} k_{h_z}(u_j)\ w_j(z) - \frac{k_{h_z}(u_i)\ w_i(z)}{2}, \tag{4.2}$$

where the vector of sample weights w must be rescaled for each z such that $\sum_{i=1}^{n} k_{h_z}(u_i)\, w_i(z) = 1$ holds for any $z \in Z$.

Similar to chapter 2, the present chapter employs a consistent Nadaraya-Watson estimator such that the kernel weights are $k_{h_z}(u_i) = K_{h_z}(u_i)/\sum_{j=1}^{n} K_{h_z}(u_j)$ with $\sum_{i=1}^{n} k_{h_z}(u_i) = 1$. The quartic (biweight) kernel with $K(u_i) = {}^{15}\!/_{16}\left(1-u_i^2\right)^2 I_{|u_i|<1}$, $u_i = z_i - z$ and $K_{h_z}(u_i) = (1/h_z)K(u_i/h_z)$ is applied. I_A denotes an indicator function which is 1 if restriction A is is fulfilled and zero otherwise. The quartic kernel assigns higher weights to observations closer to z (smaller u_i), lower weights to observations further away from z (larger u_i) and zero weight if an observation is outside the bandwidth. Fan and Gijbels (1992) have shown that adaptive local smoothers generally yield good results and avoid the well-known boundary effect. The bandwidth parameter h_z is here chosen inversely to the local data density f_z.

It has further been argued in chapter 2 that comparing concentration indices of dichotomous outcomes with varying prevalences may be misleading. According to Wagstaff (2005, 2011), the bounds of concentration indices for binary variables depend inversely on the mean μ, $|C| \leq 1-\mu$. To allow comparisons of inequalities throughout the support of z, one may adapt Wagstaff's (2005) approach such that $W(z) = C(z)/(1-\mu(z))$ is the varying Wagstaff index. $W(z)$ is then rescaled to a $(-1;1)$ interval irrespective of $\mu(z)$. This index has the advantage that it compensates changes in the concentration index caused by pure changes of the underlying prevalence. See chapter 2 for a more technical introduction and further computational details.

4.4 Results

4.4.1 Results from the 2009 sample

The results in table 4.1 for the full sample show that mean age is lower in the male than in the female sample. The average net equivalent household income is higher for men than for women in the full and the restricted sample. Women have a considerably lower prevalence of ever and current smoking than men. The homogeneous Wagstaff indices demonstrate that current smoking is significantly concentrated among the worse-off for both sexes in the full and the (restricted) ever-smoker sample. Conversely, former smoking is concentrated among the better-off in all samples. Ever-smoking is significantly concentrated among the lower income households in the male sample while no significant concentration is observed in the female sample. This is also reflected by the average income which is lower for men among the male ever-smokers compared with the full samples. According to table 4.1, the prevalence of current smoking among the ever-smokers is higher for women than for men. In contrast

Table 4.1: Descriptive smoking statistics (2009 sample)

	male			female		
	mean	W	σ_W	mean	W	σ_W
full sample	$n = 151,151$			$n = 165,825$		
age	47.03^a			48.93^a		
income	1,597.32^a	0.2896^{b*}	0.0046^c	1,493.36^a	0.2807^{b*}	0.0040^c
ever-smokers	56.26%	–0.0834*	0.0071	37.36%	–0.0039	0.0059
current smokers	31.33%	–0.1687*	0.0063	22.12%	–0.1108*	0.0065
former smokers	24.93%	0.0843*	0.0067	15.24%	0.1408*	0.0073
ever-smokers	$n = 84,745$			$n = 60,131$		
age	48.64^a			45.36^a		
income	1,541.43^a	0.2862^{*b}	0.0062^c	1,493.60^a	0.2850^{*b}	0.0069^c
current smokers	55.69%	–0.1814*	0.0089	59.21%	–0.2027*	0.0101
former smokers	44.31%	0.1814*	0.0095	40.79%	0.2027*	0.0121

Mean net equivalent household incomes and prevalences of current, ever and former smoking for the male and female full and restricted samples from the 2009 microcensus. *) significant at the 99 percent level; ^{a}) mean of variable; ^{b}) Gini index (without Wagstaff's correction) for income; ^{c}) standard error for varying concentration (here Gini) index

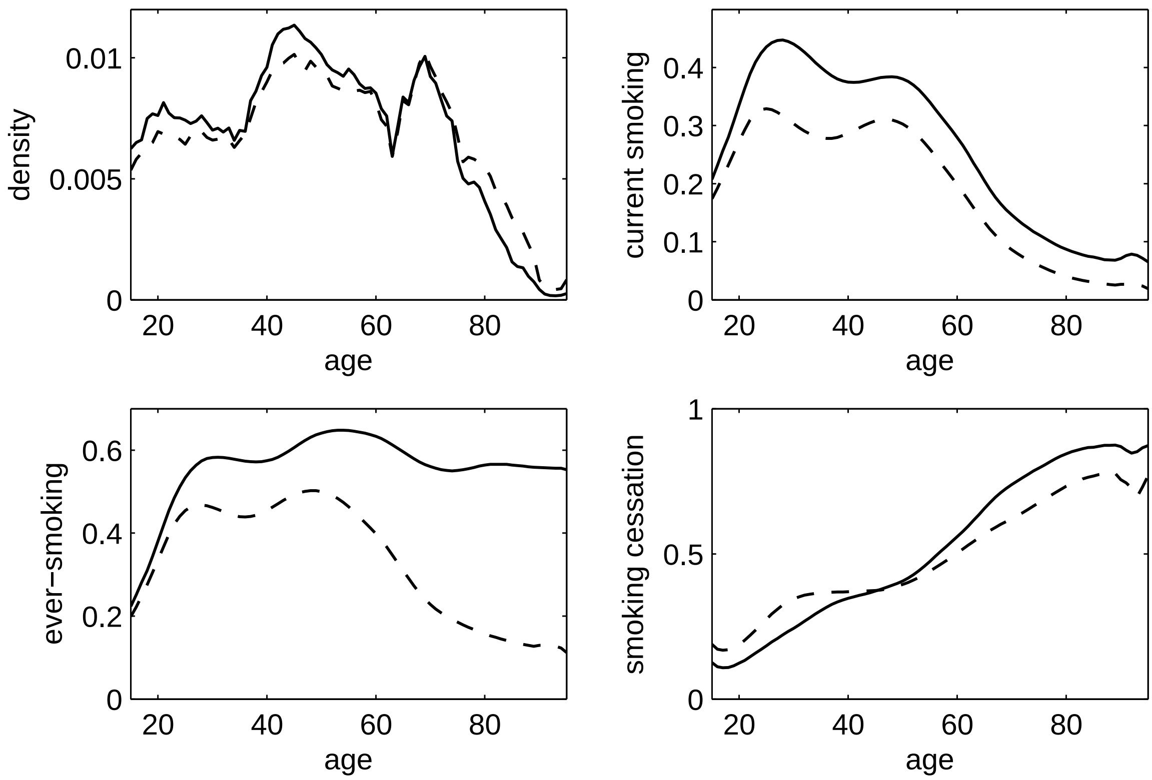


Figure 4.1: Descriptive figures (2009 sample)
Empirical density of age (upper left) and prevalence of current smoking (upper right), ever-smoking (bottom left) and smoking cessation among ever-smokers (bottom right) for the male (solid lines) and female (dashed lines) samples from the 2009 microcensus.

to the full sample, ever-smoking women are, on average, younger than ever-smoking men. The Wagstaff index for former smoking in table 4.1 is exactly the negative of the Wagstaff index for current smoking in the ever-smoking subsample for both sexes. Note that the index must fulfill the so-called mirror condition $w_{\text{current}} = -w_{\text{former}}$ because ever-smokers can only be current or former smokers (Erreygers, 2009; Wagstaff, 2011).

The data density plot in figure 4.1 presents the distribution of age in the male and female sample and corresponds with the population pyramid for Germany. The lower density for those younger than 40 corresponds to the often discussed problem of an aging society (von Weizsäcker, 1996). The data density estimate is higher for the male than for the female sample for those younger than 68 and higher for the female sample for those who are older. This result is in line with the higher average age for females found in table 4.1 and may be explained by their higher life expectancy.

The age-specific prevalence of current smoking in figure 4.1 shows similar patterns for men and women. It starts around 21 percent for males and 17 percent for females aged 15 and is higher for the male sample in all age groups. It increases rapidly for both sexes peaking around age 29 for males and 25 for females. The prevalence varies around approximately 40 percent for men and 30 percent for women in the age groups between 30 and 50. It then decreases until age 80 and flattens out around 7 percent for elder men and 3 percent for elder women.

The ever-smoking curve for females in figure 4.1 is shaped similarly to the current smoking curve but at an approximately 10 percent higher level for adult women. In contrast, the ever-smoking curve for men differs from that for current smoking. In all cohorts older than 26, almost 60 percent of the men are ever-smokers while the prevalence of current smoking is lower for those older than 50. The current and ever-smoking curves for the female sample both suggest lower prevalences for those older than 50. One may read this as a cohort effect suggesting that smoking became common among females only among younger birth cohorts (i.e. born after 1950).

The share of ever-smokers who quit smoking starts around 13 percent for the youngest males and 19 percent for the youngest females and is lowest for the 17 years old. The smoking cessation rate then increases until age 80 where it flattens out. The curves intersect at age 46. Among ever-smokers older than 46, females are less likely to have quit smoking then males.

The varying Wagstaff indices in figure 4.2 suggest a concentration of current smoking among the worse-off. The index for men is negative for all age groups and significant for those younger than 80. In contrast, women exhibit an insignificant concentration among the better-off older than 74. While the concentration is lower for younger males and strengthens among the 20 to 30 years old, it varies around some constant value until age 50 in the female

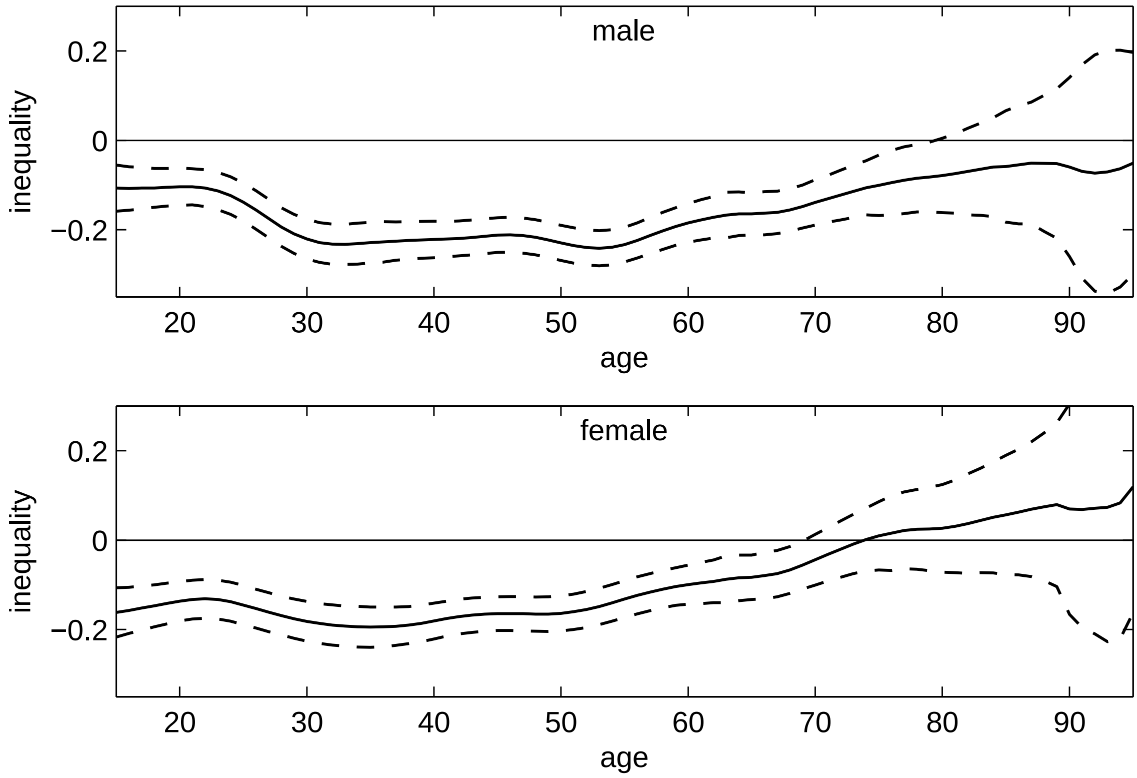


Figure 4.2: Age-specific inequality of current smoking (2009 sample)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples from the 2009 microcensus.

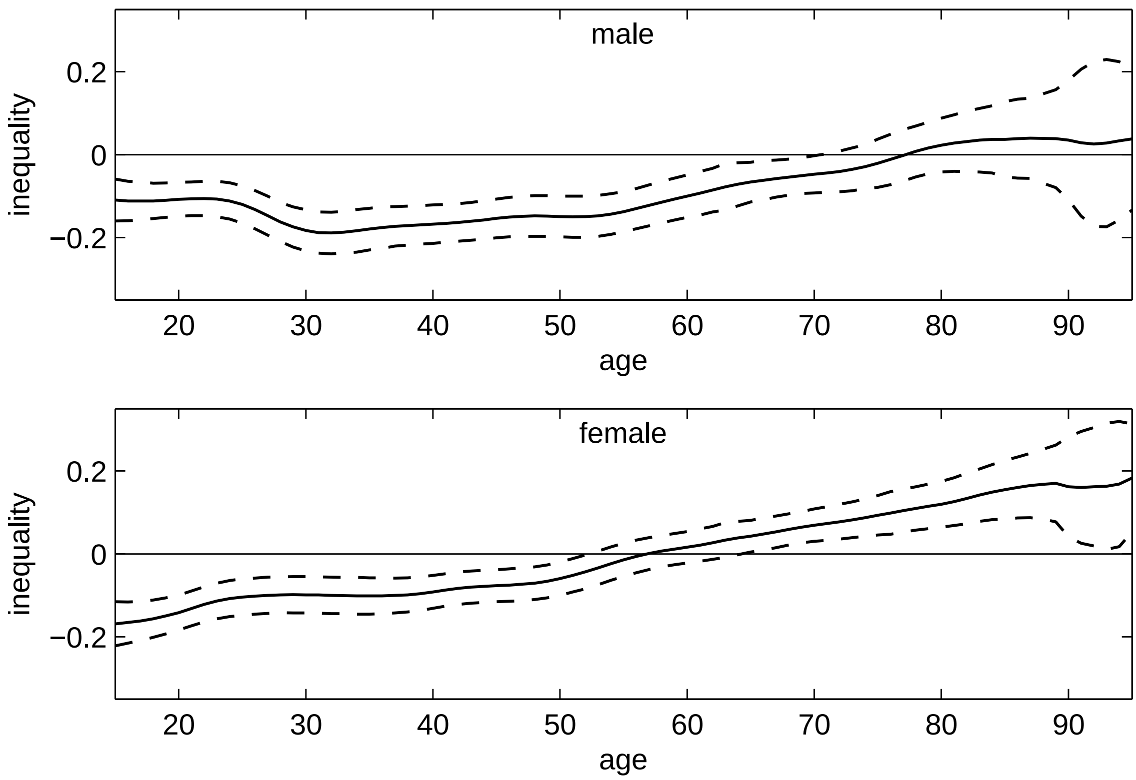


Figure 4.3: Age-specific inequality of ever-smoking (2009 sample)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples from the 2009 microcensus.

sample. One who is interested in gender aspects may wish to compare the indices for both sexes. The confidence intervals for the indices for the male and female samples do not overlap for those aged between 52 and 59. This may be considered as a significantly stronger concentration of current smoking among lower incomes in the male compared with the female sample in this age group.

Similar to the graphs in figure 4.2, the curves for ever-smoking in figure 4.3 show a positive trend with increasing age for men over 30 and all women. Ever-smokers are (at the 5 percent level) significantly concentrated among the worse-off for males younger than 70 and females younger than 52. In contrast to the male sample, a significant concentration of ever-smokers among the higher incomes is observed for females older than 65. One may read this as a cohort effect indicating that the risk of smoking ever in life shifted from the rich towards the poor during the twentieth century; the change was more pronounced in the female sample. The concentration of ever-smoking among lower-income adolescents is stronger in the female than in the male sample for those younger than 23. For the older cohorts, women have a weaker concentration among the poor or, where the index is positive, a stronger concentration among the better-off than men. The confidence intervals for the estimates from the male and female samples do not overlap for those aged between 50 and 75. This indicates that the concentration of ever-smokers in lower-income households is significantly stronger among males in these age groups. Comparing the results in figure 4.3 with those in figure 4.2, one may note that the curves for current and ever-smoking follow similar patterns but at a somewhat higher level for the latter.

Figure 4.4 presents the age-specific inequality indices for smoking cessation estimated from the restricted ever-smokers subsample. The graph suggests that higher income ever-smokers are more likely to quit smoking than those in lower income households. This is in line with the stronger concentration of current smoking compared with ever-smoking among the worse-off. The concentration among the better-off is significant at the 5 percent level for males between 23 and 72 and females between 22 and 74. Comparing the varying Wagstaff indices for men and women yields no significant gender differences.

4.4.2 Results from the 2005 sample[1]

Comparing the results in table 4.2 with table 4.1, it can be observed that individuals in the 2005 sample are, on average, somewhat younger than in the 2009 sample. The mean net equivalent household income increased while the prevalences of ever-smoking, current smoking and former smoking decreased marginally between 2005 and 2009. The observed income inequality as well as the income-related inequalities of smoking behavior do not differ significantly.

[1] The 2005 sample is described in section 2.3.

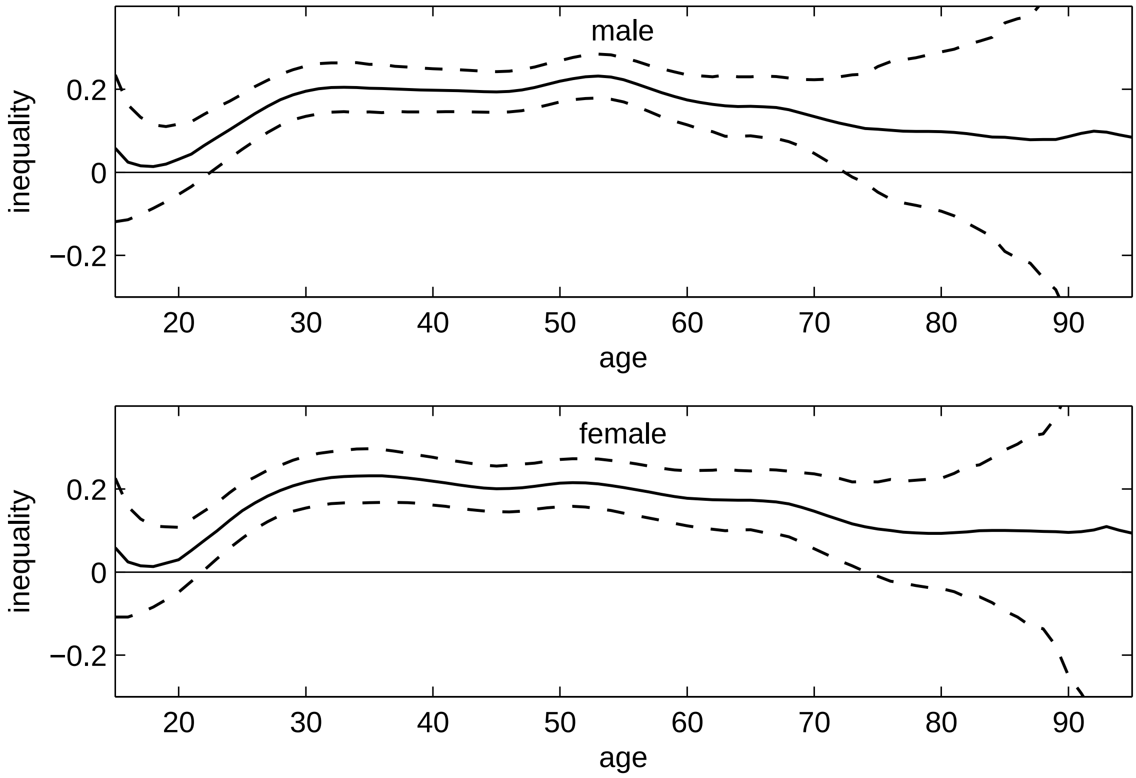


Figure 4.4: Age-specific inequality of smoking cessation (2009 sample)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples from the 2009 microcensus.

Table 4.2: Descriptive smoking statistics (2005 sample)

	male			female		
	prevalence	W	σ_W	prevalence	W	σ_W
full sample	$n = 155,251$			$n = 169,286$		
age	45.66[a]			47.79[a]		
income	1,471.89[a]	0.2910[b]*	0.0042[c]	1,383.81[a]	0.2824[b]*	0.0040[c]
ever-smokers	57.02%	–0.0783*	0.0070	37.73%	0.0093	0.0059
current smokers	32.85%	–0.1670*	0.0062	23.49%	–0.1090*	0.0064
former smokers	24.17%	0.0963*	0.0066	14.24%	0.1426*	0.0075
ever-smokers	$n = 88,336$			$n = 61,614$		
age	47.09[a]			43.38[a]		
income	1,424.61[a]	0.2880[b]*	0.0054[c]	1,385.68[a]	0.2902[b]*	0.0067[c]
current smokers	57,62%	–0.1873*	0.0093	62.26%	–0.2022*	0.0130
former smokers	42.38%	0.1873*	0.0083	37.74%	0.2022*	0.0100

Mean net equivalent household incomes and prevalences of current, ever and former smoking for the male and female full and restricted samples from the 2005 microcensus. *) significant at the 99 percent level; [a]) mean of variable; [b]) Gini index (without Wagstaff's correction) for income; [c]) standard error for varying concentration (here Gini) index

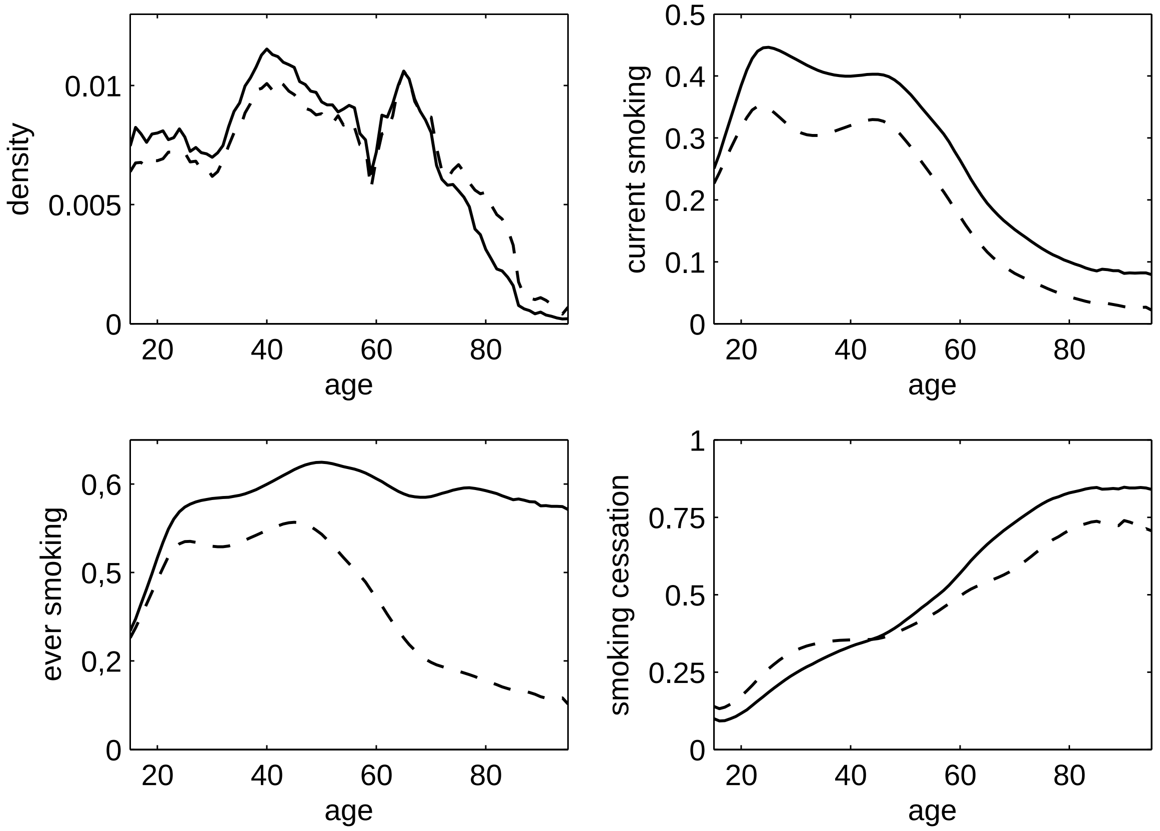


Figure 4.5: Descriptive figures (2005 sample)
Empirical density (upper left) and prevalence of current smoking (upper right), ever-smoking (bottom left) and smoking cessation among ever-smokers (bottom right) for the male (solid lines) and female (dashed lines) samples from the 2005 microcensus.

The changes in age and income between 2005 and 2009 are similar in the restricted sample. Similarly, the composition of the ever-smokers sample in terms of current and former smokers differ only marginally. Income inequality and income-related inequality of former and current smoking among ever-smokers do not differ significantly. The age-specific descriptive statistics for the 2005 microcensus sample in figure 4.5 are similar to those for the 2009 sample in figure 4.1. It can be observed that the graphs for ever-smoking and smoking cessation in the 2005 sample seem to be shifted somewhat to the left. This indicates that similar values were observed for the same birth cohorts in both samples.

Regarding the age-specific inequalities in smoking, it is found that the results for current

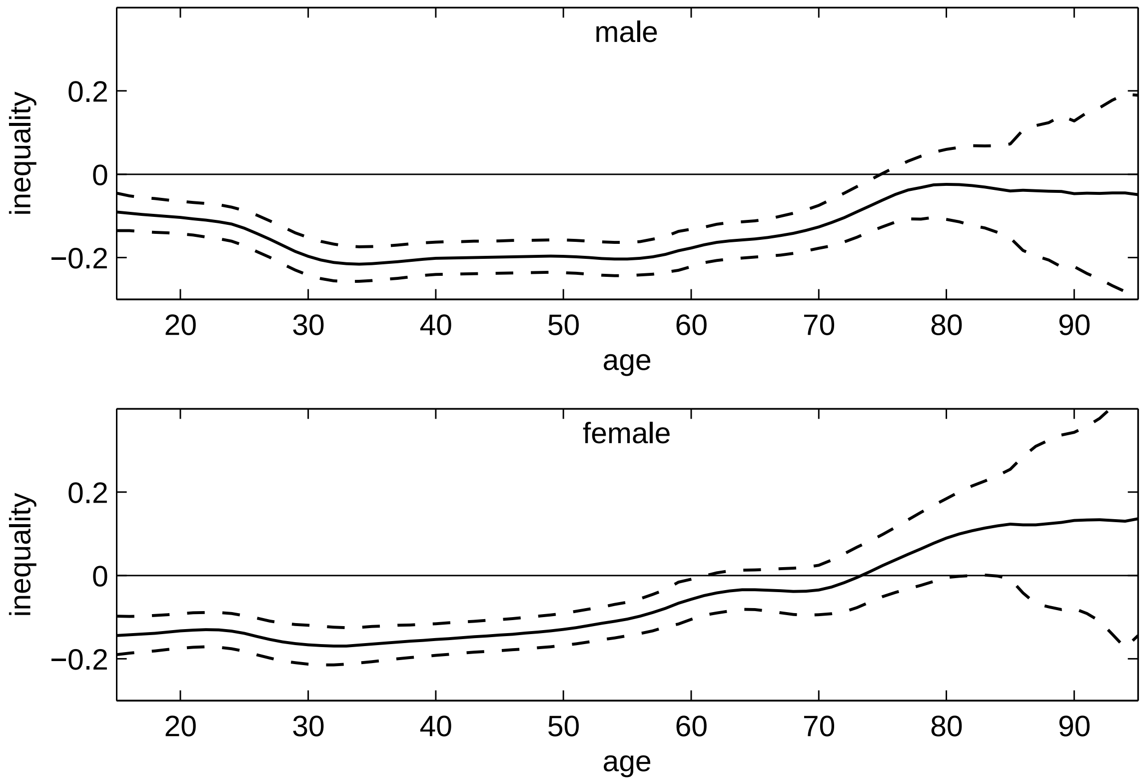


Figure 4.6: Age-specific inequality of current smoking (2005 sample)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples from the 2005 microcensus.

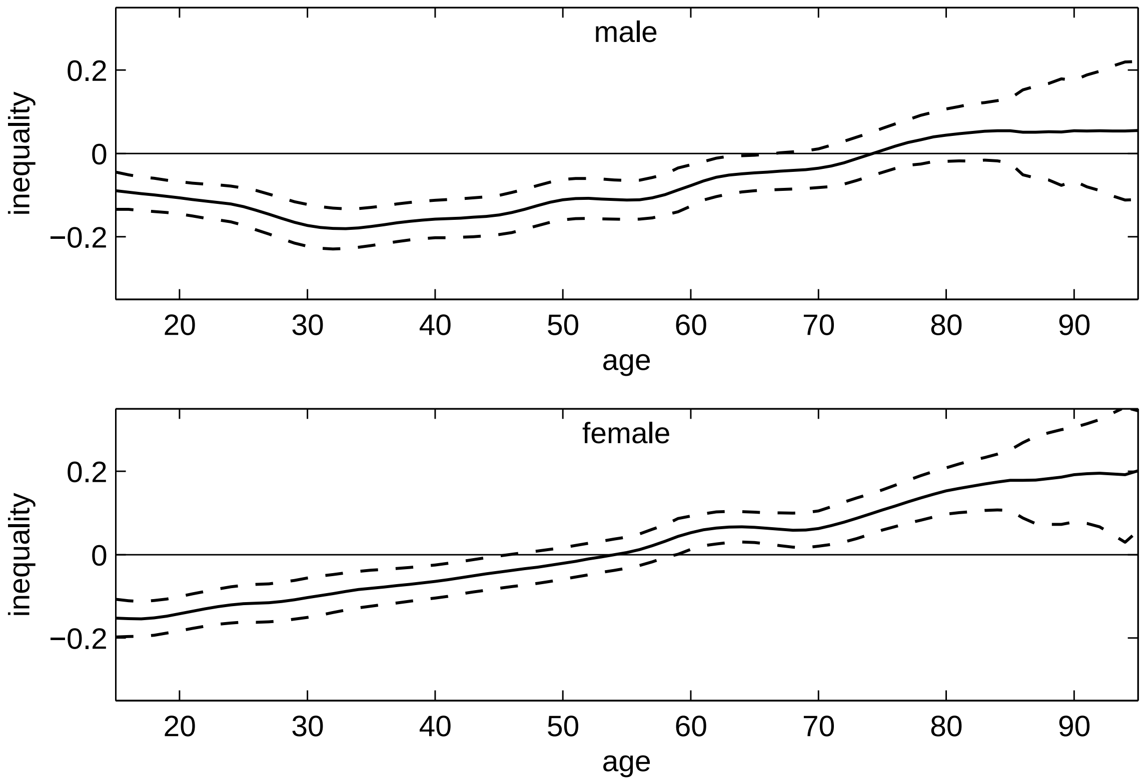


Figure 4.7: Age-specific inequality of ever-smoking (2005 sample)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples from the 2005 microcensus.

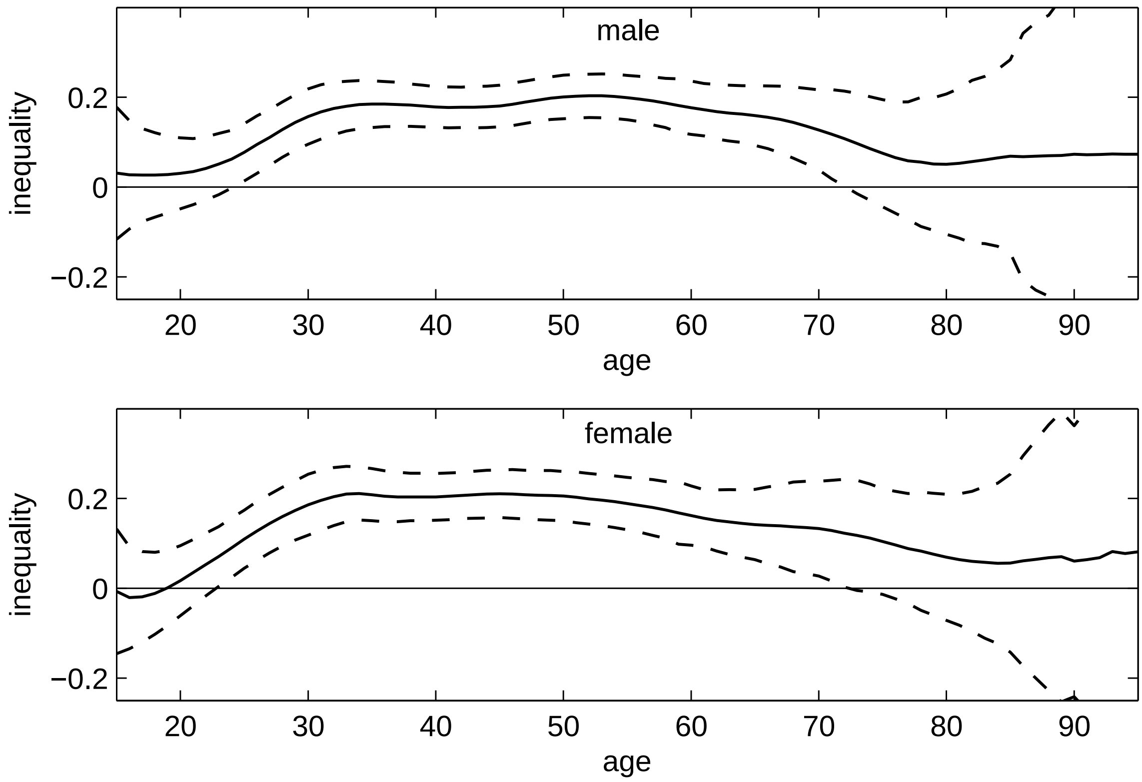


Figure 4.8: Age-specific inequality of smoking cessation (2005 sample)
Age-specific inequality indices (solid lines) with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples from the 2005 microcensus.

smoking from the 2005 sample in figure 4.6 do not differ significantly from those obtained from the 2009 sample in figure 4.2. Comparing the results for ever-smoking, one may agree that the overall patterns of the graphs in figure 4.3 and 4.7 are fairly similar. One may note that the turnover point where the income-related gradient shifts from a pro-rich towards a pro-poor distribution is approximately in the same birth cohort for both samples. For the age-specific inequality of smoking cessation, figures 4.4 and 4.8 do not suggest significant changes between the 2005 and 2009 samples.

4.4.3 A note on nonresponses

Although the similarities between the results obtained from the 2009 and 2005 microcensus samples suggest that the results are fairly reliable, one may be interested in the age-specific income-related concentration of nonresponses to the voluntary smoking module. According to the results in figure 4.9, males and females in households with lower net equivalent household incomes were less likely to answer the questions concerning smoking behavior. Figure 4.10 demonstrates that the relation between income and nonresponse was considerably weaker in the sample from the 2005 microcensus.

Given that the results from both samples are fairly similar, one may agree that considerable selection biases seem to play a minor role here. However, it should be mentioned that an income-related concentration of nonresponse may lead, under the assumption that the relation between income rank and smoking behavior also holds among the nonrespondents, to an underestimation of the income-related gradient to the detriment of economically deprived households in smoking.

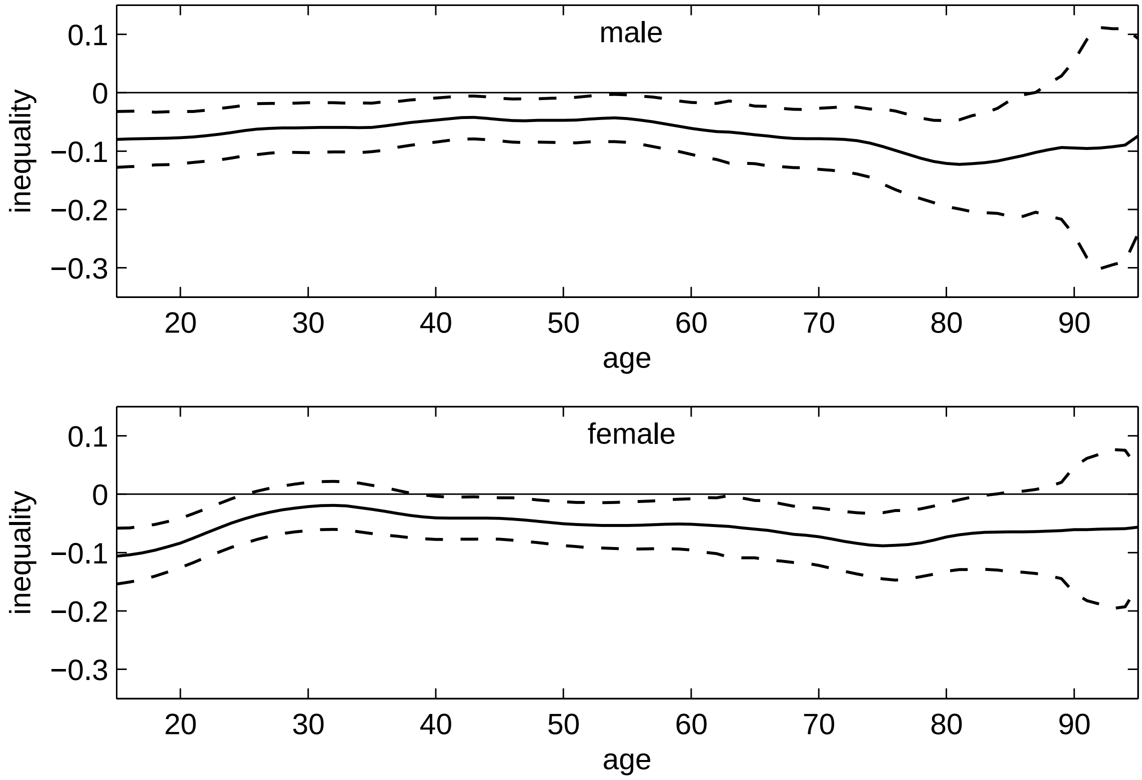


Figure 4.9: Nonresponse to smoking module (2009 sample)
Age-specific inequality indices (solid lines) for with 95 percent confidence intervals (dashed lines) for nonresponse to the smoking module in male (upper) and female (bottom) samples from the 2009 microcensus.

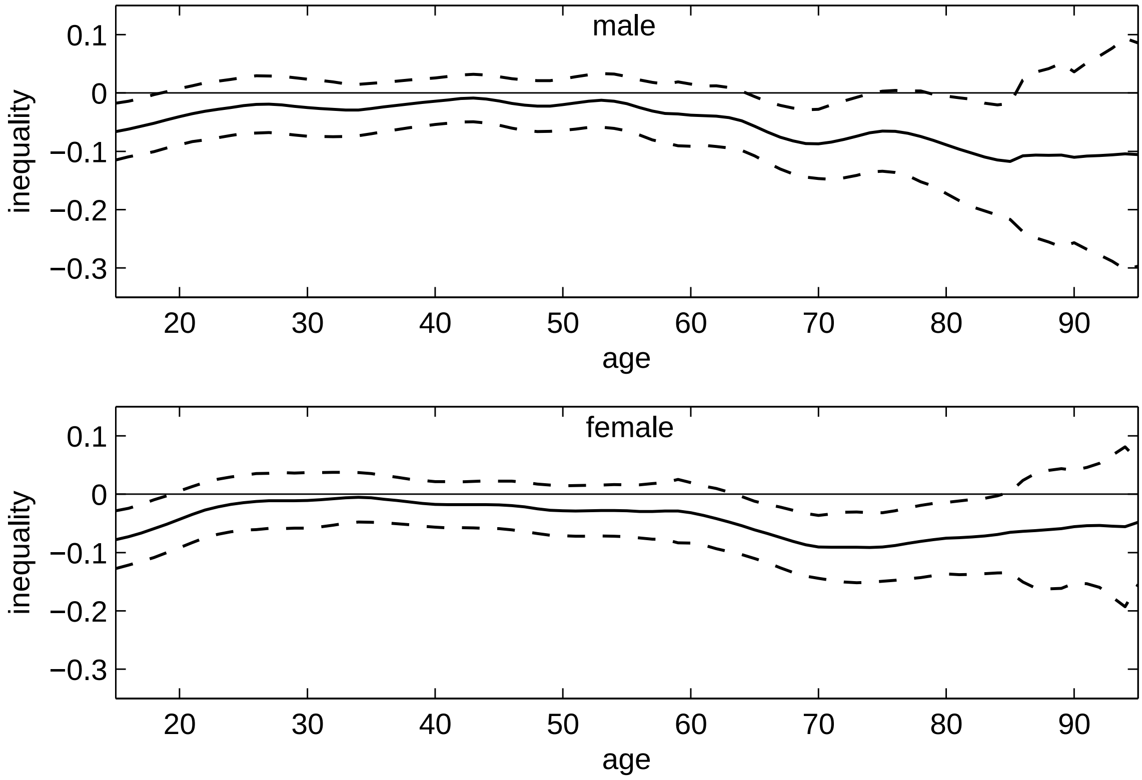


Figure 4.10: Nonresponse to smoking module (2005 sample)
Age-specific inequality indices (solid lines) for with 95 percent confidence intervals (dashed lines) for nonresponse to the smoking module in male (upper) and female (bottom) samples from the 2005 microcensus.

4.5 Discussion

The present chapter applied the varying inequality index based on the concept of Gini-type concentration indices introduced in chapter 2 to data on smoking behavior from the German microcensus. It has been found that the income-related gradients vary considerably with age. The results suggest that a homogeneous index would neither have revealed the lower concentration of current and ever-smokers among adolescents in lower income households nor the pro-rich distribution of current and ever-smoking among the elderly. In contrast to Richter and Leppin (2007), a significant gradient related to household income for adolescents of both sexes is observed here. Smoking cessation exhibits no significant income-related gradients for the youngest. One may consider two reasons for this: first, the estimates for the varying Wagstaff indices are close to zero among the youngest. Second, only few stopped smoking among the youngest. It is important to mention here that such low prevalence rates increase the uncertainty and hence widen the confidence intervals.

Bauer et al. (2007) stress the importance of gender-specific policies to reduce smoking efficiently. The microcensus data reveal an increasing gap between men and women in the prevalence of current and ever-smoking with increasing age (in other words decreasing for later birth cohorts). This is in line with the common result that gender differences reduced during the twentieth century. For all age groups, however, men still exhibit a higher prevalence of current and ever-smoking than women. Somewhat surprisingly, smoking cessation is more common among male than female ever-smokers. Female ever-smokers are more likely to quit smoking only until age 46. In older age groups, women who ever smoked are less likely to be former smokers than men. As Bauer et al. (2007) could not attribute such gender differences to differences in socioeconomic position and endowments, the authors assume them to be mainly a behavioral effect which may be explained with gender roles. The socioeconomic gradients for current and ever-smoking computed from the microcensus data suggest similar patterns for males and females in terms of income-related inequalities among those older than 30. Despite the similarities, the concentration of current and ever-smoking among the worse-off is significantly weaker in the female sample at ages 52-59 for current and 50-75 for ever-smoking.

The above results for the inequality in ever-smoking suggest that its income-related concentration shifted from the higher to the lower incomes during the twentieth century. The varying Wagstaff indices presented in figures 4.3 and 4.7 change from pro-rich (positive index) to pro-poor (negative index) distributions around the 1930 birth cohort in the male sample (age 78 in figure 4.3 and age 75 in figure 4.7) and around the 1950 birth cohort in the female sample (age 57 in figure 4.3 and age 55 in figure 4.7). One may object measuring cohort effects in socioeconomic gradients via household income and argue that income may vary over the life course while e.g. education could be considered as a durable asset. However, Schulze and Mons (2006) found similar results for the educational dimension of inequalities in smoking. They identify a change from the better to the less educated between the 1921-30 and 1931-40 birth cohorts for men and between the 1931-40 and 1941-50 birth cohorts for women. Comparing age-specific smoking prevalences for different educational levels in the underlying data yields similar results.

Analyses of health inequalities over the life course based on cross-sectional data may be subject to certain biases. It is widely agreed that life expectancy is lower among the deprived (see e.g. Balia and Jones, 2008). As smoking is related to severe diseases and premature mortality (Genuneit et al., 2006; Kamholz, 2004; Peto et al., 2000; Slama, 2008), smoking-related mortality may be considered as a possible confounder. Comparing the results with the overall mortality rates from the Human Mortality Database (Human Mortality Database, 2012), one may agree that mortality is unlikely to bias the results considerably before age 70. Another issue may be a potential bias through bad health selection into early retirement. However, this should lead to opposite results at least for ever-smokers. One may agree that, in contrast to the observed results, the adverse health consequences should accumulate among the worse-off and lead to a concentration of smoking among them. Current and ever-smoking are, however, pro rich for the oldest.

4.6 Conclusions

The distributions of current and ever-smoking among adolescents and young adults is significantly pro-poor. Smoking cessation, in contrast, is more common among the better-off ever-smokers (note that in the full sample, former smoking is pro-poor among adolescents and pro-rich among adults). The results suggest that anti-smoking policies should aim at adolescents and young adults in lower-income households.

One may conclude from the results that the smoking epidemic first started among the better-off. The smoking prevalence increased somewhat later among the economically deprived while smoking apparently became less common among the better-off. The patterns are quite similar for males and females; the results suggest that the developments started earlier for males than for females, though. It seems that the smoking epidemic proceeds similarly for both sexes but with some delay for females.

The results suggest that most males and females, if ever, start smoking in adolescence or early adulthood. Smoking prevalence is high among the younger adults and individuals apparently stop in mid-life. Sundmacher (2012) has shown that smoking cessation is closely related to diagnoses of related diseases. One may speculate that such diseases rarely occur before mid-life and consider this as a possible explanation for the patterns of the current smoking and smoking cessation curves. One may further speculate that the higher rates of smoking cessation among younger females compared with males may be a maternity effect as the average number of dependent infants is particularly high in households with 20 to 40 year old women.

Chapter 5

Using an alternative measure of socioeconomic status: Community deprivation-related health inequalities[1]

5.1 Introduction

Population is usually ranked by some individual or household level variable in health inequality analyses using (rank-based) concentration indices. For instance, Wagstaff et al. (2003) rank children in Vietnam by per capita household consumption; and Jones and López Nicolás (2006), van Doorslaer et al. (2004) and van Doorslaer and Koolman (2004), as well as chapters 2, 3 and 4 use net equivalent household income to facilitate the socioeconomic ranking in industrialized countries.

The question whether, and to what extent, health is influenced by neighborhood or community deprivation has become an increasingly important issue in epidemiology and health inequality analyses (see e.g. Maier et al., 2011; Kuznetsov et al., 2011, 2012). Noble et al.

[1] An enhanced version of this chapter is planned to be published as a joint article with Andreas Mielck and Werner Maier.

(2006) describe five Indices of Multiple Deprivation for the United Kingdom: the English Indices of Deprivation 2000, the English Indices of Deprivation 2004, the Welsh Index of Multiple Deprivation 2000, the Northern Ireland Measures of Multiple Deprivation 2001 and the Scottish Indices of Deprivation 2003. Maier et al. (2011) were the first to use the key principles described in Noble et al. (2006) to propose a Bavarian Index of Multiple Deprivation. This index has been applied to Bavarian health data repeatedly (Kuznetsov et al., 2011, 2012) and was recently extended to a German Index of Multiple Deprivation.

After analyzing age-specific variations in income-related health inequalities in the previous chapters, the present chapter addresses two related questions:

1. To what extent can age-specific community deprivation-related health inequalities be observed?

2. How do income-related health inequalities vary with relative community deprivation?

The varying inequality index introduced in chapter 2 is applied to data on obesity, hypertension and diabetes drawn from the pooled sample of the 2002 and 2006 Health Care Access Panel which were already used in chapter 3. The German Index of Multiple Deprivation [2] is used to determine the community deprivation.

5.2 Data and variables

Data for the empirical analysis are drawn from the TNS Health Care Access Panel provided by Kantar Health (formerly TNS Infratest Healthcare). To obtain sufficiently large numbers of observations throughout all age groups and regions, the 2002 and 2006 waves of the Health Care Access Panel are pooled for the analysis (see Potthoff et al., 2004 and section 3.2 for further descriptions of the data).

[2] I would like to acknowledge that Werner Maier kindly provided me the within-domain rankings and the theory-based Index of Multiple Deprivation ready for use.

Of the 117,167 individuals (48,574 households) included in the pooled sample, 75,122 individuals (29,421 households) participated in 2002 and 60,555 individuals (28,828 households) in 2006. 18,510 individuals (8,718 households) participated in both surveys. 28,390 individuals younger than 20 were excluded as chronic conditions rarely affect younger individuals and a meaningful interpretation of the body mass index is problematic for children and adolescents. 1,176 observations had to be removed owing to missing data on income.[3] Another 7,991 observations had to be dropped owing to unsuitable regional codes (*Gemeindekennziffer*). The sample for the empirical analysis eventually comprises 79,610 individuals (41,767 females and 37,843 males) in 43,652 households.

The modified OECD equivalence scale is applied to compute net equivalent household income as a measure of socioeconomic status in terms of income-related health inequalities. The first health outcome is obesity defined as a body mass index (BMI) over 30 (see WHO, 2009). The body mass index is computed as body weight in kg divided by the squared body height in meters from self-reported anthropometric data. The second health outcome is hypertension within the preceding twelve months and the third is diabetes mellitus. Note that the survey does not allow a unique distinction between type one and type two diabetes. One may, however, consider type two diabetes as age related and influenceable through lifestyles and health behavior (see Harati et al., 2010; Puska, 2010) and type one diabetes as mainly genetic and thus likely to be equally distributed across socioeconomic groups. As in chapter 3, diabetes is analyzed regardless of its type or insulin dependency.

[3] Up to this point, data preparation and the resulting sample are exactly the same as in chapter 3.

5.3 Methods

5.3.1 Building the Index of Multiple Deprivation

The Index of Multiple Deprivation has been developed to rank communities or administrative areas by a set of weighted deprivation domains. The domains included in such an index have to be relevant in the context of deprivation and suitable indicator variables for each domain must be available (Noble et al., 2006). Such indicators must be

> *"(1) 'domain specific' and appropriate for the purpose (as direct as possible measures for that form of deprivation), (2) measuring major features of that deprivation (not conditions just experienced by a very small number of people or areas), (3) up-to-date, (4) capable of being updated on a regular basis, (5) statistically robust, and (6) available for the whole of the country in question at a small-area level in a consistent form"* (Noble et al., 2006, p. 176).

Based on these requirements, Maier et al. (2011) define seven domains for the Bavarian and the German Index of Multiple Deprivation. The variables and domains included in the Bavarian Index of Multiple Deprivation are listed in table 5.1. The German Index of Multiple Deprivation is computed similarly here; with one exception, though. The crime rates included in the security domain for the Bavarian index are not included in the German Index of Multiple Deprivation.[4] The indicator variables are standardized using the *z*-score transformation. Noble et al. (2006) propose to perform a maximum likelihood factor analysis to combine the chosen indicators to univariate domains. However, as the maximum number of indicators per domain is two, techniques such as factor analyses would not contribute to the present analysis. Maier et al. (2011) argue that such techniques are only required when combining at least three indicators into one domain. Where only one single indicator is included, this variable is the only factor and therefore the main factor. Where two indicators are combined to one domain, it is

[4]Crime rates at the community level for overall Germany are currently not available for the computation.

Table 5.1: Domains of the German Index of Multiple Deprivation

domain	indicators	theory	weights from factor analysis
personal income	total income per tax payer	25%	39.51%
employment	unemployment rate per employable population (age 15 - 65)	25%	39.68%
education	rate of unskilled workers per employees s.t. social insurance contributions	15%	–15.37%
communal income	communal gross revenues surplus/shortfall and debts per citizen	15%	7.76%
social capital	migration rates (communal level) and voter participation	10%	32.19%
environment	share of sealed surface (*commercial, industrial and traffic*)	5%	–4.10%
personal security	average number of traffic accidents per citizen crime rates per citizen[a]	5%	0.33%

Source (*excluding the right column*): Maier et al. (2011, table 3). The left column shows the seven domains included in Maier et al.'s Index of Multiple Deprivation. The center column lists the indicators included in each domain. The two right hand columns present the theory-based weights proposed by Maier et al. (2011) and weights obtained through factor analysis performed on the 2002 and 2006 sample from the Health Care Access Panel. [a]) Note that the crime rates are only included in the Bavarian index because they were not available at the community level for overall Germany when the German Index of Multiple Deprivation used in the present analysis was generated.

technically impossible to identify a main factor. Both variables would be equally weighted because algorithms such as principal component or factor analysis would consider them as equipollent.

The domain variables computed in the previous step are then used to rank communities in ascending order by deprivation. These rankings could be used as domain deprivation indices for domain-specific analyses. One who is interested in an Index of Multiple Deprivation, however, now needs to combine the domains into one index. This requires a judgment to what extent more deprivation in one domain may be compensated by less deprivation in another domain. It has been argued that z-score transformation and untransformed rankings within domains may both lead to such cancellations (Noble et al., 2006). Maier et al. (2011) follow Noble et al. (2006) who argue that exponential transformation of the domain ranks may tackle this problem. Denoting ρ as the communal rank within a domain, the best-off commune has $\rho = 1/n$ (least deprived) and the worst-off has $\rho = 1$ (most deprived). The transformed within-domain rank is then

$$D = -\delta \ln\left\{1 - \rho\left(1 - \exp\left\{-\frac{100}{23}\right\}\right)\right\}. \tag{5.1}$$

Noble et al. (2006) and Maier et al. (2011) both advise to choose the factor $\delta = 23$. They point out that this choice is fully arbitrary but has the advantage that all domain variables D are transformed to a $(0;100)$ scale, where only the most deprived decile obtains values larger than 50.

The separate domains then have to be combined to a single Index of Multiple Deprivation. Note that it is impossible to do this without deciding on a weighting scheme; even simply summing up the domains would correspond to arbitrarily chosen equal weights. Weights should be carefully chosen, though; for example based on theories derived from the literature or on empirical methods. An alternative may be a somewhat arbitrary weighting scheme with respect to policy relevance. Either way, the weighting scheme will have considerable impact on the final index and researchers should be aware of the implications of whatever weighting scheme they choose (see Noble et al., 2006, who discuss this in detail). The weighting scheme for the British indices has been derived from the literature and put a focus on income and employment (see Noble et al., 2006). Maier et al. (2011) adopt the British scheme and similarly assign the highest weights to these two domains. The lowest weights are assigned to the environment and security domains. The weights are presented in table 5.1.

Noble et al. (2006) mention that empirically driven weighting schemes obtained through factor analysis may be an alternative to the above described theory-based approach. To avoid cancellation effects, this chapter uses the domain scores D obtained from equation (5.1) to obtain empirically justified weights through factor analysis. For each domain, D is z-score transformed to facilitate identification of a main factor through maximum likelihood estimation techniques. The weights obtained from this empirically driven approach are given the right column in table 5.1.

5.3.2 Measuring inequality

Age-specific and community deprivation-specific income-related health inequalities, as well as age-specific community deprivation-related health inequalities, are measured using the semiparametric extension of the concentration index introduced in chapter 2. The index was derived from the well-known concentration index C which has become a common technique to measure socioeconomic gradients in health (see e.g. Wagstaff et al., 1991; Wagstaff and van Doorslaer, 2000; van Doorslaer et al., 2000; Kakwani et al., 1997). C stems from the concentration curve where the cumulative share of some health variable y is plotted against the cumulative share of the population ranked by socioeconomic status. The curve lies below (above) the line of equality (45° line), if y concentrates among the better-off (worse-off). Measuring twice the area between the concentration curve and the 45° line, C is bound in the $(-1;1)$ interval and becomes positive (negative), if the concentration curve lies below (above) the line of equality. Where no inequality is observed, i.e. all individuals have the same level of y regardless of their socioeconomic status, the concentration curve is exactly the line of equality and C is zero. Konings et al. (2010) and O'Donnell et al. (2008) provide intuitive introductions to the concept of concentration curves and indices.

The estimation of homogeneous (overall-sample) concentration indices has been discussed elaborately elsewhere (see e.g. van Doorslaer et al., 1997, 2004; van Doorslaer and Koolman, 2004; Kakwani et al., 1997; Konings et al., 2010; O'Donnell et al., 2008; Wagstaff et al., 1991; Wagstaff and van Doorslaer, 2000; Wagstaff et al., 2003, to mention a few). The present chapter applies the varying inequality index introduced in chapter 2. To obtain the concentration index as a function of some other parameter $z \in Z \subset \mathbb{R}$, the convenient regression approach (Kakwani et al., 1997; Wagstaff et al., 2003; Wildman, 2003) is combined with the varying coefficient model (Hastie and Tibshirani, 1993; Li et al., 2002). The formula for the convenient regression is

$$2\frac{\sigma_r^2(z)}{\mu(z)}y = \beta_0(z) + \beta_1(z)r(z) + \varepsilon, \tag{5.2}$$

where $C(z) = \beta_1(z)$ with $z \in Z$ and $\mu(z)$ is the z-specific mean of y. $C(z)$ and $\mu(z)$ can both be obtained through nonparametric regression. The fractional rank has to be computed locally as

$$r_i(z) = \sum_{j=1}^{i} k_{h_z}(u_j) - \frac{k_{h_z}(u_i)}{2}, \tag{5.3}$$

where $z \in Z$ and $u_i = z_i - z$. Individuals i must be sorted in ascending order by socioeconomic status and $k_{h_z}(u_i) = K_{h_z}(u_i)\left[\sum_{j1}^{n} K_{h_z}(u_j)\right]^{-1}$ are the kernel weights with $\sum_{i=1}^{n} k_{h_z}(u_i) = 1$ and n being the number of observations. Computing the rank variable locally using equation (5.3) assures that the locally computed mean and variance of the local rank variable are (asymptotically) $1/2$ and $1/12$, respectively, for any given $z \in Z$ (see also chapter 2). The varying concentration index is estimated using a Nadaraya-Watson estimator with a quartic kernel function. The kernel function assigns higher weights to observations closer to z, lower weights to observations further away from z and zero weights if an observation is outside the bandwidth h_z.

The local bandwidth parameter h_z is chosen inversely to the local density f_z to find an optimal balance between bias and uncertainty. Where z denotes age, the bandwidth is chosen as $h_z = 1.06\,\hat{\sigma}_z n^{-0.2}\,\hat{f}_z^{-0.3}$ with $\hat{f}_z$ being the empirical density at a given $z \in Z$ and $\hat{\sigma}_z$ the standard error of z obtained from the data (see chapter 2). Where z is the communal deprivation rank, it has been found that choosing the bandwidth as $h_z = 1.06\,\hat{\sigma}_z n^{-0.1}\,\hat{f}_z^{-0.3}$ performs well.

As the bounds of the concentration index for binary variables depend inversely on the mean, comparisons particularly across age groups may lead to misleading results. For the age-specific analyses, the varying concentration index is therefore corrected using Wagstaff's (2005) formula such that $W(z) = C(z)/(1-\mu(z))$; see section 2.2.3 for details. Pointwise confidence intervals are reported using the variance approximation described in sections 2.2.4 and 2.2.5.

5.3.3 Testing rank sensitivity

Wagstaff and Watanabe (2003) derive a straightforward method to compute and statistically test the differences between concentration indices based on different socioeconomic status variables (for further descriptions and applications see also Lindelow, 2006; O'Donnell et al., 2008).

Let two different socioeconomic status variables yield two different vectors of fractional ranks, r_1 and r_2. The homogeneous concentration index using the first ranking variable r_1 is

$$C_1 = \frac{2}{n\mu}\sum_{i=1}^{n} y_i r_{1i} - 1, \tag{5.4}$$

C_2 is computed analogously. Wagstaff and Watanabe (2003) have shown that the difference between both concentration indices can be written as

$$\begin{aligned}
\Delta C = C_2 - C_1 &= \frac{2}{n\mu}\sum_{i=1}^{n} y_i r_{2i} - 1 - \left[\frac{2}{n\mu}\sum_{i=1}^{n} y_i r_{1i} - 1\right] \\
&= \frac{2}{n\mu}\sum_{i=1}^{n} y_i (r_{2i} - r_{1i}) \\
&= \frac{2}{n\mu}\sum_{i=1}^{n} y_i \Delta r_i,
\end{aligned} \tag{5.5}$$

where $\Delta r_i = r_2 - r_1$ is individual i's change of the fractional rank. Wagstaff and Watanabe (2003) and Lindelow (2006) have shown that ΔC can be computed via the convenient regression approach,

$$\frac{2\sigma^2_{\Delta r}}{\mu} y = \beta_0 + \beta_1 \Delta r + \varepsilon, \tag{5.6}$$

where $\sigma^2_{\Delta r}$ is the variance of Δr. The advantage of this straightforward method is that one may compute the standard error $\sigma_{\Delta C}$ for $\Delta C = \beta_1$ using equation (5.6) (Lindelow, 2006; O'Donnell et al., 2008; Wagstaff and Watanabe, 2003). The sample variability of μ is taken into account here following Wagstaff and Watanabe (2003) and O'Donnell et al. (2008), respectively.

Wagstaff and Watanabe (2003) stress that $\Delta C = 0$ does not necessarily imply that socioeconomic ranks do not differ between two socioeconomic status variables, i.e. that $\Delta r_i = 0$ for all i. By measuring the covariance between the change of the rank variable, however, it would indicate that the health variable does not vary with Δr. In other words, it shows to what extent the decision how to measure socioeconomic status influences the measured inequality.

5.3.4 Indirect standardization

Figure 5.1 demonstrates that health status is strongly related to age.[5] The oldest have the highest prevalence rates but, on average, the lowest incomes. This is unproblematic when per-

[5] The unstandardized age-specific prevalences in figure 5.1 look somewhat similar to those in figure 3.1. The 7,991 observations dropped owing to unsuitable regional codes apparently have not caused considerable changes in the age-specific prevalences compared to the sample used in chapter 3.

forming age-specific analyses. In other cases, however, one has to account for the age-specific distribution of the outcome variables. For the community deprivation-specific analyses here, neglecting the above mentioned age dependency of health and income may evoke the so-called student or pension effects (see Islam et al., 2010, and section 3.5). The pension effect describes that older people after retirement have worse health outcomes and lower incomes compared with younger individuals in their working lifes which may cause an artificial concentration of bad health among the poor. In contrast, the student effect describes that young individuals in the beginning of their working lifes are, on average, healthy but have low incomes. This may cause an underestimation of health disadvantages among the worse-off.

A common approach to tackle this problem is to statistically remove the purely age-related health effects by means of indirect standardization (van Doorslaer et al., 2000; O'Donnell et al., 2008; Wagstaff and van Doorslaer, 2000). The procedure is as follows: Let x be the matrix of demographic indicators to be used for standardization (age groups here) and the predicted risk at a given x be $\hat{y} = Pr(y = 1|x)$. The formula for the indirect standardization is then

$$y_i^* = y_i - (\hat{y}_i - \tilde{\mu}), \tag{5.7}$$

where y_i^* is the x-standardized health variable for the i-th individual (O'Donnell et al., 2008). The mean $\tilde{\mu}$ of the age predicted health variable $\hat{y}$ is included to assure that the overall sample mean of the standardized variable equals the overall sample mean of the unstandardized variable, $\mu^* = \mu$. Note that O'Donnell et al. (2008) present the formula using μ in place of $\tilde{\mu}$ such that $y_i^* = y_i - (\hat{y}_i - \mu)$. Numerically, however, $\mu = \tilde{\mu}$ does not necessarily hold where $\hat{y}_i$ is obtained through nonlinear regression.

For the present analysis, data were stratified by sex and individuals were then grouped by age into five year intervals, each indicated by a dummy variable. The age-predicted risks of obesity, hypertension and diabetes were obtained through logistic regression using maximum likelihood estimation. The age-standardized prevalences were then computed using equation

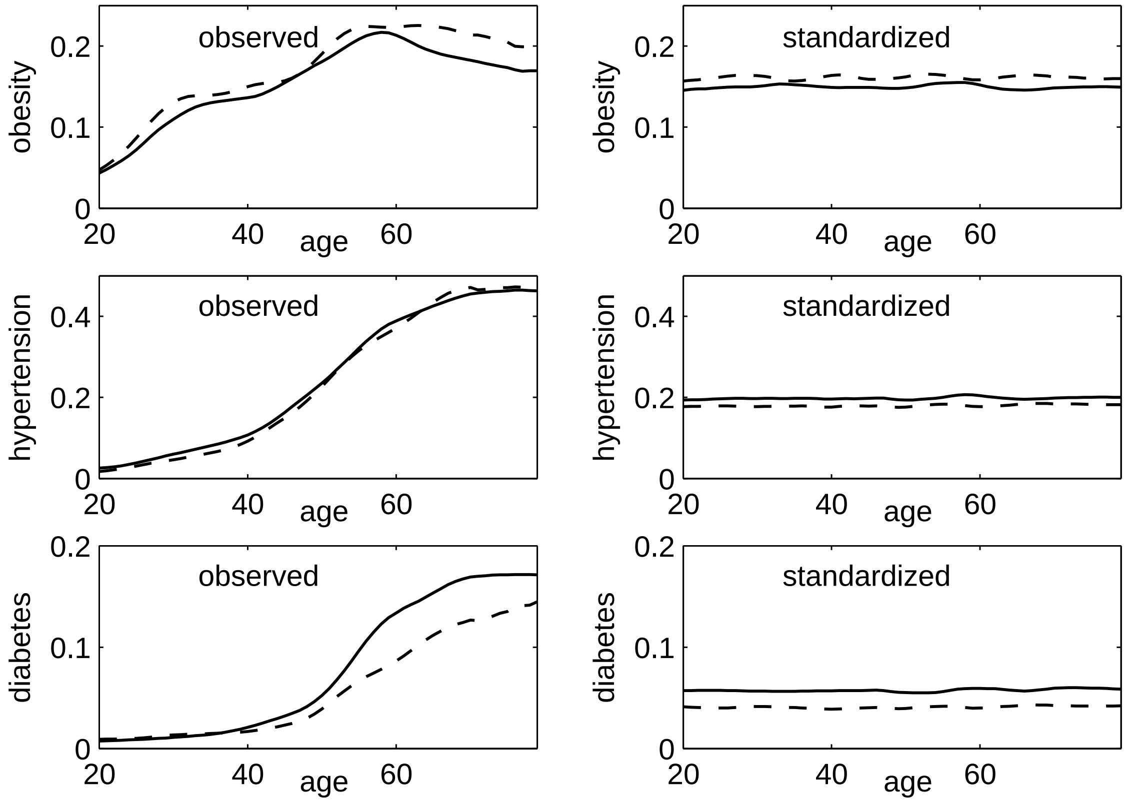


Figure 5.1: Age-specific and standardized prevalences (pooled sample 2002, 2006)
Age-specific (left) and age-standardized (right) prevalences of obesity (top), hypertension (middle) and diabetes (bottom) for the male (solid lines) and female (dashed lines) samples from the 2002 and 2006 Health Care Access Panel.

(5.7).

The unstandardized and the standardized age-specific prevalences are presented in figure 5.1. While the observed age-specific prevalences in the left column vary with age (similar results were found in figure 3.1), the standardized prevalences are almost constant across all age groups. The latter will be used in this chapter wherever overall-sample or community deprivation-specific statistics are computed. The unstandardized values will be used for all age-specific analyses.

5.4 Results

5.4.1 Comparing the two community deprivation indices

Comparing the weighting schemes for the theory-based and the factor analysis based Indices of Multiple Deprivation in table 5.1, one may agree that the two approaches produce considerably different results. Both approaches highlight individual income and unemployment rates, however, the results from factor analysis are much more focused on these two domains.[6] Both together account for approximately 79.2 percent in the weighting scheme derived from factor analysis while accounting for only 50 percent in the weighting scheme derived from the literature. Further, the proportion of the social capital domain is 30.2 percent in the factor analysis based weighting scheme while theory suggests a weight of only 10 percent. One may consider the higher weight for the social capital domain to be plausible as high migration rates and low voter participation may be consequences of lacking opportunities and sustained dissatisfaction with an area or commune. In contrast to the theory-based approach, the factor analysis based index does not consider higher rates of unskilled workers as deprivation. By assigning a negative weight to this domain, the index indicates the opposite: Unskilled workers are more likely to be employed in areas with higher incomes, higher voter participation, lower unemployment rates and less emigration (and vice versa). According to the theory-based approach, communal revenues are among the more important domains, however, factor analysis assigns approximately half the weight (7.8 percent). Both approaches give the lowest (absolute) weight to personal security.

Despite the different weighting schemes, the rank correlation coefficients in table 5.2 suggest a high correlation between the theory-based and the factor analysis based deprivation indices. Although education has a comparably high weight in the weighting scheme derived from theory, the corresponding rank correlation coefficient of 0.0368 is comparably low. The

[6] Principal component analysis using Spearman's rank correlation matrix yields a similar index. Its correlation coefficient with the factor analysis based index is 0.9874; the corresponding rank correlation coefficient is 0.9738.

Table 5.2: Correlation matrix of overall deprivation ranks (pooled sample 2002, 2006)

	Index of Multiple Deprivation		domain deprivation rank by					
	theory based	factor analysis	income	employment	education	communal revenue	social capital	environment
factor analysis	0.7983							
deprivation domain								
income	0.6440	0.8856						
employment	0.7979	0.9124	0.6714					
education	0.0368	–0.3219	–0.2524	–0.3341				
comm. revenue	0.6238	0.3221	0.2246	0.3237	–0.0339			
social capital	0.5922	0.6539	0.5556	0.5130	0.0235	0.1141		
environment	0.2650	0.0486	–0.1936	0.2732	0.1395	0.0863	–0.795	
security	0.0975	0.0242	–0.0561	0.0732	0.0865	–0.0826	0.0975	0.2199

Spearman's rank correlation coefficients for the multiple and single domain deprivation ranks from the 2002 and 2006 Health Care Access Panel. All rank correlation coefficients are highly significant (p-values < 0.001).

low coefficients of deprivation in terms of environment (0.0486) and security (0.0242) with the factor analysis based deprivation index reflect their low absolute weights. As prefigured by the weighting schemes derived from factor analysis in table 5.1, the deprivation rank in the education domain is negatively correlated with the ranks in the per capita income, employment and communal revenue domains. Somewhat surprisingly, only education is negatively correlated with the factor analysis based index despite the negative weight also assigned to the environment domain. The negative correlation between per capita income deprivation and environmental deprivation likely reflects higher average incomes in urban compared with rural areas. Similarly, the negative correlation between the environment and social capital domain ranks indicates migration flows towards more urbanized (higher income) areas. These interpretations are confirmed by the rank correlation between community population size and the deprivation ranks in the income domain (negative sign), environment domain (positive sign) and social capital domains (negative sign) observed in the data.

5.4.2 Overall-sample estimates and rank sensitivity

Before presenting the results for deprivation-related inequalities, it should be emphasized that higher deprivation ranks often represent worse-off communes in the literature. To retain comparability of the results with the other chapters, however, all Gini and concentration indices (both varying and homogeneous) are computed using deprivation in descending order: similar to the computation of income-related concentration indices where a higher rank indicates a better-off household, higher ranks for deprivation-related inequalities represent better-off communes.

The results in table 5.3 suggest that mean income is higher in the male than in the female sample. For both sexes, the average income is similar to that in table 3.1. The income-related concentration indices are all highly significant. The Gini-index for income is similar to that in table 3.1 and roughly corresponds to the estimates from the 2005 and 2009 microcensus data

Table 5.3: Rank sensitivity: household income and Index of Multiple Deprivation derived from the literature (pooled sample 2002, 2006)

	prevalence	individual income		Index of Multiple Deprivation[a]		sensitivity	
		C	σ_C	C	σ_C	ΔC	$\sigma_{\Delta C}$
male				$n = 37,843$			
income	1,463.51[b]	0.2996[c]*	0.0018	0.0408*	0.0020	–0.2588*	0.0020
obesity	14.98%	–0.0592*	0.0069	–0.0347*	0.0071	0.0245*	0.0094
hypertension	19.76%	–0.0238*	0.0056	–0.0267*	0.0056	0.0029	0.0074
diabetes	5.71%	-0.0858*	0.0118	–0.0503*	0.0117	0.0355†	0.0153
female				$n = 41,767$			
income	1,396.60[b]	0.2969[c]*	0.0017	0.0411*	0.0018	–0.2558*	0.0019
obesity	16.12%	–0.1209*	0.0062	–0.0395*	0.0064	0814*	0.0084
hypertension	17.89%	–0.0675*	0.0055	–0.0366*	0.0056	0.0309*	0.0074
diabetes	4.05%	–0.1550*	0.0133	–0.0597*	0.0135	0.0953*	0.0177

Means and concentration indices of income and age-standardized health variables from the 2002 and 2006 Health Care Access Panel. †) significant at the 95 percent level; *) significant at the 99 percent level. [a]) Index of Multiple Deprivation derived from the literature (Maier et al., 2011). [b]) Mean of net equivalent household income; [c]) Gini index of net equivalent household income (see section 2.4 and table 4.1, respectively). Alike the results for income-related inequalities, the concentration indices with respect to the Indices of Multiple Deprivation are all highly significant. The deprivation-related concentration of income is considerably lower than the Gini index of income for both sexes. While the income-related concentration indices differ significantly between the male and female samples, the deprivation-related estimates in table 5.3 do not. In summary, males exhibit weaker income-related concentration of diseases among the worse-off than females while no gender-specific differences are found for inequalities with respect to the Index of Multiple Deprivation based on weights derived from the literature.

The right two columns in table 5.3 demonstrate that the estimates for the concentration indices change considerably when using community deprivation in place of equivalent income. These changes are highly statistically significant ($p < 0.01$) for income, obesity and hypertension in the female sample. In the male sample, differences are significant at the 99 percent level for income and at the 95 percent level for obesity and diabetes. To summarize, the results indicate that the rank sensitivity seems more pronounced in the female than in the male sample for the diseases and vice versa for income inequality.

Table 5.4 presents the inequalities of income and the health outcomes with respect to the theory-based and the factor analysis based deprivation indices. All concentration indices computed on the basis of the deprivation index weighted by factor analysis are highly significant. Assuming that communal per capita income is to some extent correlated with equivalent household income, one may agree that the higher community deprivation-related income inequality observed when using the index derived from the factor analysis based weighting scheme likely stems from the higher emphasis put on income deprivation here. In line with these results, the concentration of hypertension and diabetes in more deprived communes is significantly stronger when using the factor analysis based deprivation index. Sex specific differences in the factor analysis based deprivation-related concentration are significant only for hypertension.

Table 5.4: Rank sensitivity, Indices of Multiple Deprivation with theory and factor analysis based weighting (pooled sample 2002, 2006)

	Index of Multiple Deprivation					
	weights from the literature		weights from factor analysis		sensitivity	
	C	σ_C	C	σ_C	ΔC	$\sigma_{\Delta C}$
male			$n = 37\,843$			
income	0.0408*	0.0020	0.0632*	0.0018	0.0224*	0.0012
obesity	–0.0347*	0.0071	–0.0362*	0.0070	–0.0015	0.0044
hypertension	–0.0267*	0.0056	–0.0334*	0.0056	–0.0068	0.0035
diabetes	–0.0503*	0.0117	–0.0632*	0.0121	–0.0129	0.0076
female			$n = 41\,767$			
income	0.0411*	0.0018	0.0612*	0.0017	0.0202*	0.0011
obesity	–0.0395*	0.0064	–0.0454*	0.0064	–0.0059	0.0041
hypertension	–0.0366*	0.0056	–0.0503*	0.0057	–0.0138*	0.0036
diabetes	–0.0597*	0.0135	–0.0798*	0.0137	–0.0200†	0.0086

Comparison of concentration indices using the weights derived from the literature and from factor analysis (see table 5.1 and Maier et al., 2011), 2002 and 2006 Health Care Access Panel
†) significant at the 95 percent level; *)significant at the 99 percent level.

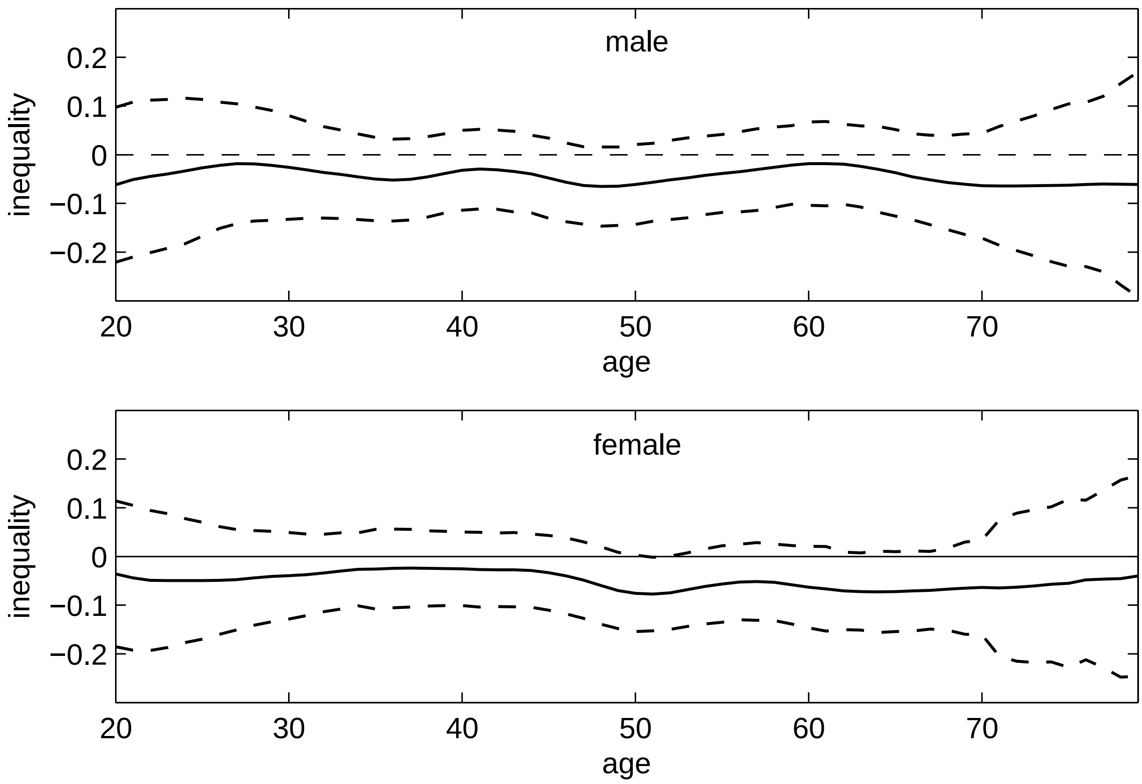


Figure 5.2: Age-specific inequality of obesity, index from theory (pooled sample 2002, 2006) Age-specific deprivation-related inequality (solid lines) with 95 percent confidence intervals (dashed lines) of obesity for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel. Deprivation index derived from the literature.

5.4.3 Age-specific variations

The rank sensitivity tests in table 5.3 demonstrate that the results based on multiple deprivation and on income differ significantly (excluding hypertension in the male sample). One may now be interested in the age-specific variations in community deprivation-related health inequalities in comparison with the income-related health inequalities analyzed in chapter 3.

Figure 5.2 indicates significant age-specific inequalities for no age group for the deprivation index computed from theory-based weights. Relevant variations in deprivation-related inequalities across age groups are not observed. Comparing this with the results in figure 3.3, it is found that the age-specific concentration of obesity is much weaker in deprived com-

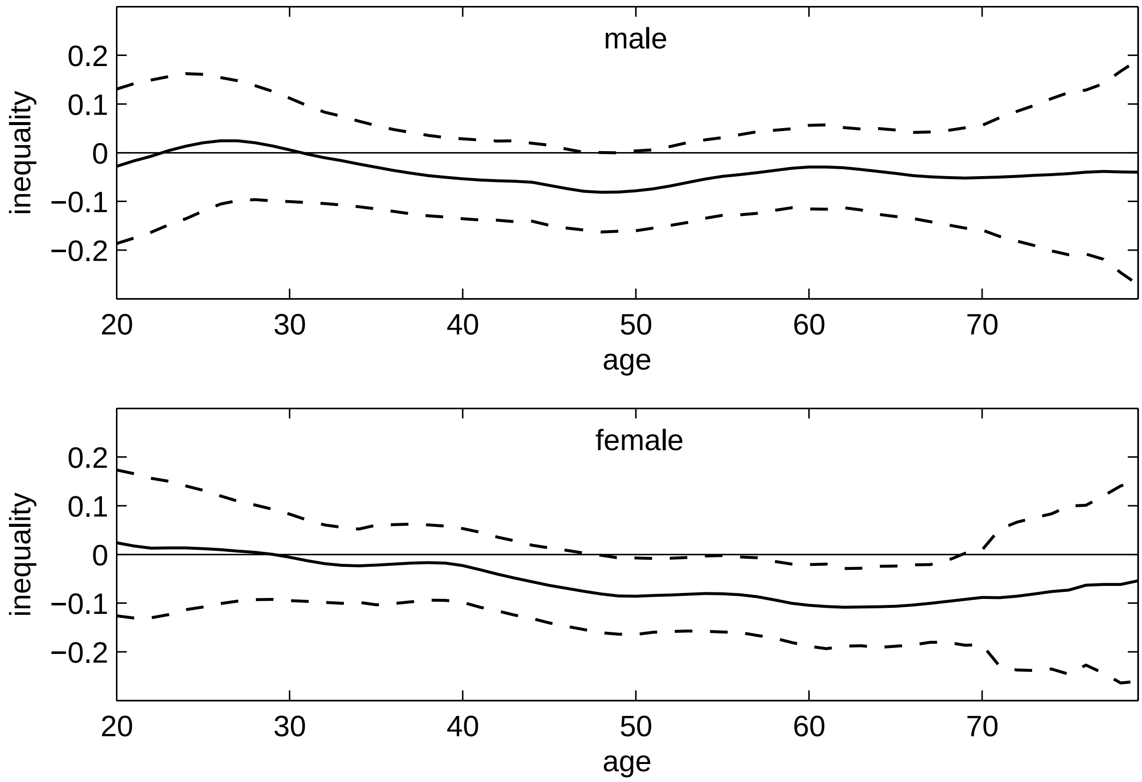


Figure 5.3: Age-specific inequality of obesity, deprivation index from factor analysis (pooled sample 2002, 2006)

Age-specific deprivation-related inequality (solid lines) with 95 percent confidence intervals (dashed lines) of obesity for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel. Deprivation index derived from factor analysis.

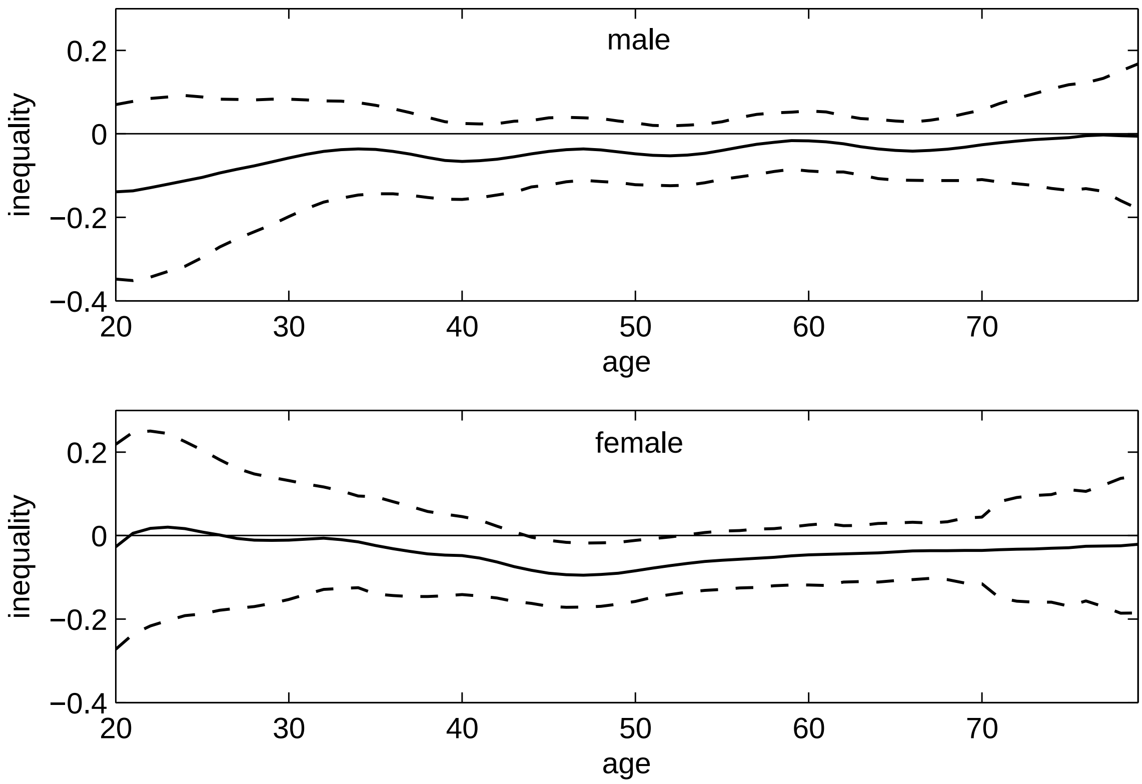


Figure 5.4: Age-specific inequality of hypertension, deprivation index from theory (pooled sample 2002, 2006)
Age-specific deprivation-related inequality (solid lines) with 95 percent confidence intervals (dashed lines) of hypertension for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel. Deprivation index derived from the literature.

munes than in lower income households. This is indicated by the lower absolute (i.e. less negative) results in figure 5.2. While the concentration of obesity among women in lower income households is statistically significant for those between 28 and 70, this only holds for the age-specific community deprivation-related inequality for the 50 years old women. Taking the results in figure 5.3 into account, one may agree that inequalities observed when using the weights derived from factor analysis as a basis are more similar to the results obtained from individual income. The shift towards a stronger health gradient observed for the female sample in figure 3.3 is, although somewhat moderated, also observed in figure 5.3.

Regarding the results for hypertension in figure 5.4, no age-specific community deprivation-

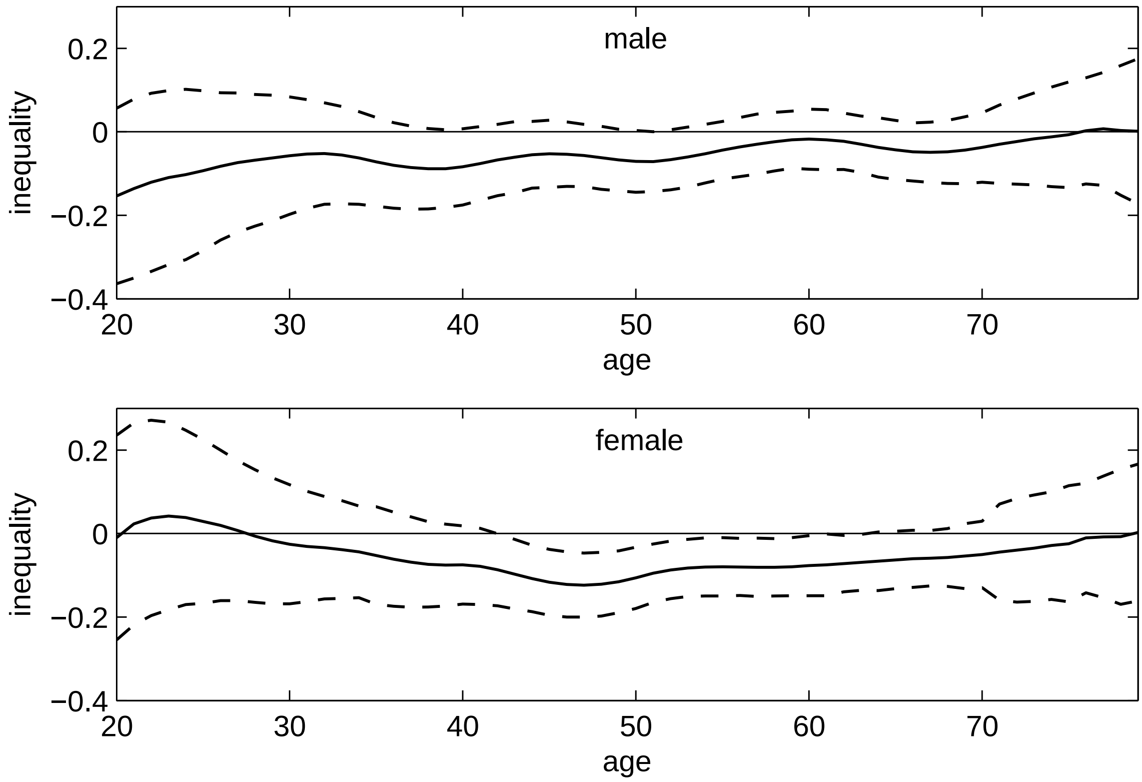


Figure 5.5: Age-specific inequality of hypertension, deprivation index from factor analysis (pooled sample 2002, 2006)

Age-specific deprivation-related inequality (solid lines) with 95 percent confidence intervals (dashed lines) of hypertension for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel. Deprivation index derived from factor analysis.

related inequalities are observed for the male sample when using the theory-based weighting scheme. Inequalities do not change considerably across age groups but are significant for the 43 to 51 years old females. Similar to the results for obesity, the observed age-specific inequalities are weaker when using community deprivation in place of household income. Note that, in contrast to the results in figure 5.4, the income-related inequalities in figure 3.4 are statistically significant for females approximately between ages 50 and 70. The results from the factor analysis based deprivation index in figure 5.5 are somewhat more similar to the income-related inequalities in figure 3.4 than those from figure 5.4. As already observed for obesity in figure 5.3, the index for hypertension in figure 5.5 yields significance in older age groups compared with the index for income-related inequalities in hypertension in figure 3.4.

Similar to the results found for obesity and hypertension, the estimates for diabetes in figure 5.6 suggest weaker health gradients when using theory-based deprivation ranks compared with the income-related health gradients in figure 3.5. While age-specific income-related inequalities are significant in the female sample for the 50 to 60 years old, no statistically significant estimates are found for community deprivation-related inequalities using the theory-based weighting scheme. This changes somewhat when using the factor analysis based weighting scheme. Figure 5.7 yields significant age-specific community deprivation-related inequalities for males around age 50 and females between ages 60 and 70. The results for the female sample yield, again, significant results in an older age group for the deprivation-related inequality compared with the income-related inequality.

5.4.4 Community deprivation-specific income-related inequalities

Tables 5.3 and 5.4 have shown that health gradients to the detriment of the socioeconomically deprived can be observed regardless of whether one uses net equivalent household income or community deprivation. Table 5.4 suggests that the results for the two Indices of Multiple

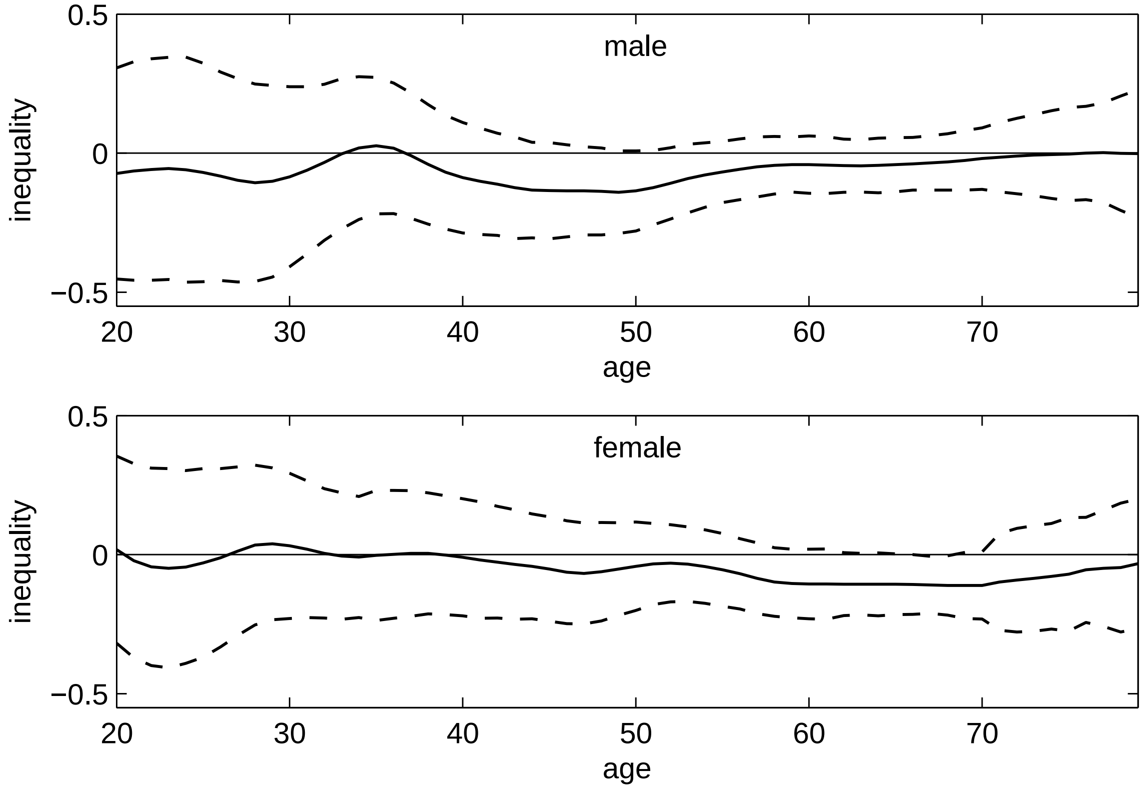


Figure 5.6: Age-specific inequality of diabetes, deprivation index from theory (pooled sample 2002, 2006)

Age-specific deprivation-related inequality (solid lines) with 95 percent confidence intervals (dashed lines) of diabetes for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel. Deprivation index derived from the literature.

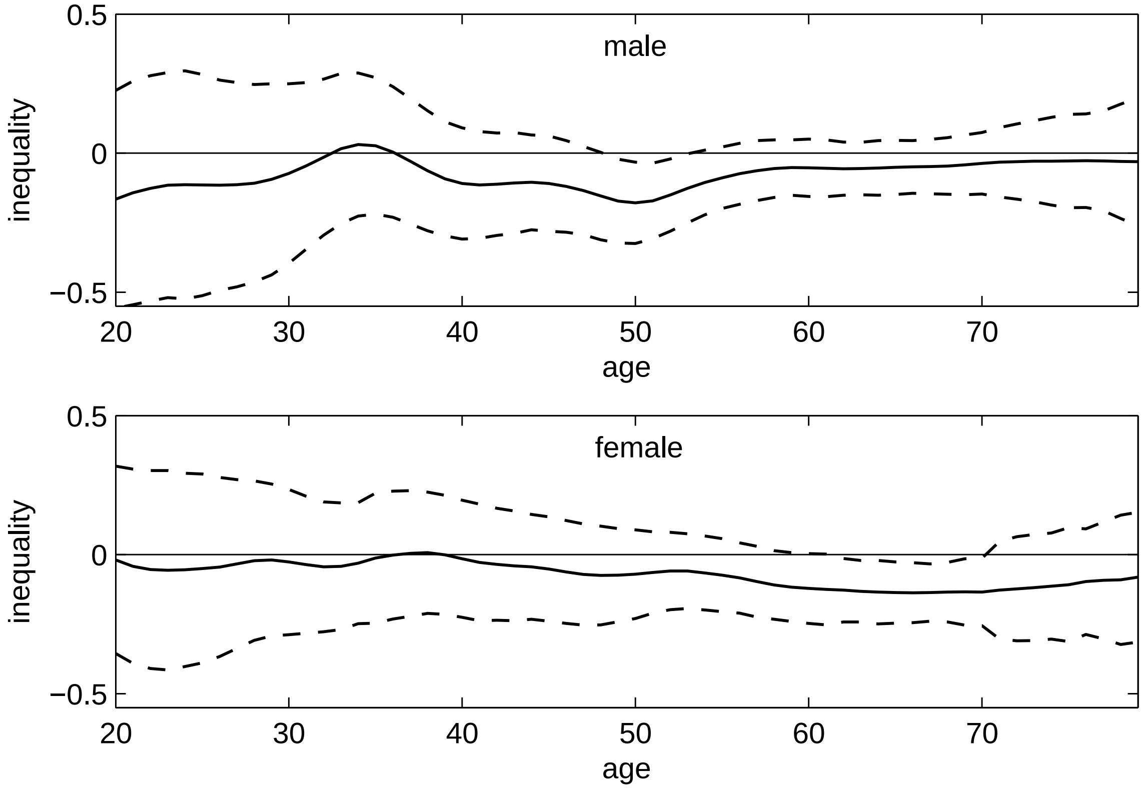


Figure 5.7: Age-specific inequality of diabetes, deprivation index from factor analysis (pooled sample 2002, 2006)

Age-specific deprivation-related inequality (solid lines) with 95 percent confidence intervals (dashed lines) of diabetes for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel. Deprivation index derived from factor analysis.

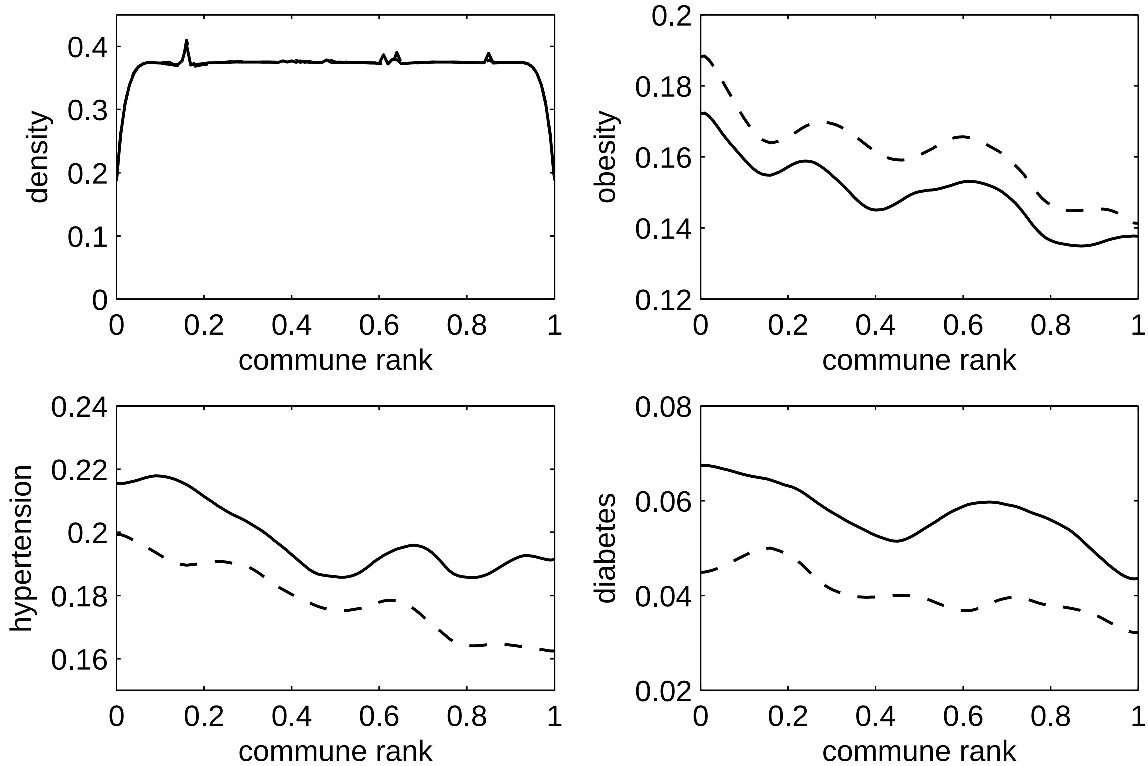


Figure 5.8: Deprivation-specific descriptive figure, deprivation index from theory (pooled sample 2002, 2006)

Empirical density (upper left) and deprivation-specific (age-standardized) prevalences of obesity (upper right), hypertension (bottom left) and diabetes (bottom right) for the male (solid lines) and female (dashed lines) samples from the 2002 and 2006 Health Care Access Panel. Deprivation index derived from the literature.

Deprivation are fairly similar at least in the male sample. The following figures demonstrate how income inequalities and income-related health inequalities vary with the communes' socioeconomic status.

The kernel density estimates in the upper left graphs of figures 5.8 and 5.9, respectively, show three noteworthy peaks in the better-off two quintiles and one in the worst-off quintile. The worst-off peak around 0.16 (theory-based, 0.22 factor analysis based) represents Berlin, Germany's largest city. A second peak is found around 0.61 (0.6 factor analysis based) which corresponds to Hamburg. The third peak found at a rank of approximately 0.64 (0.63 factor

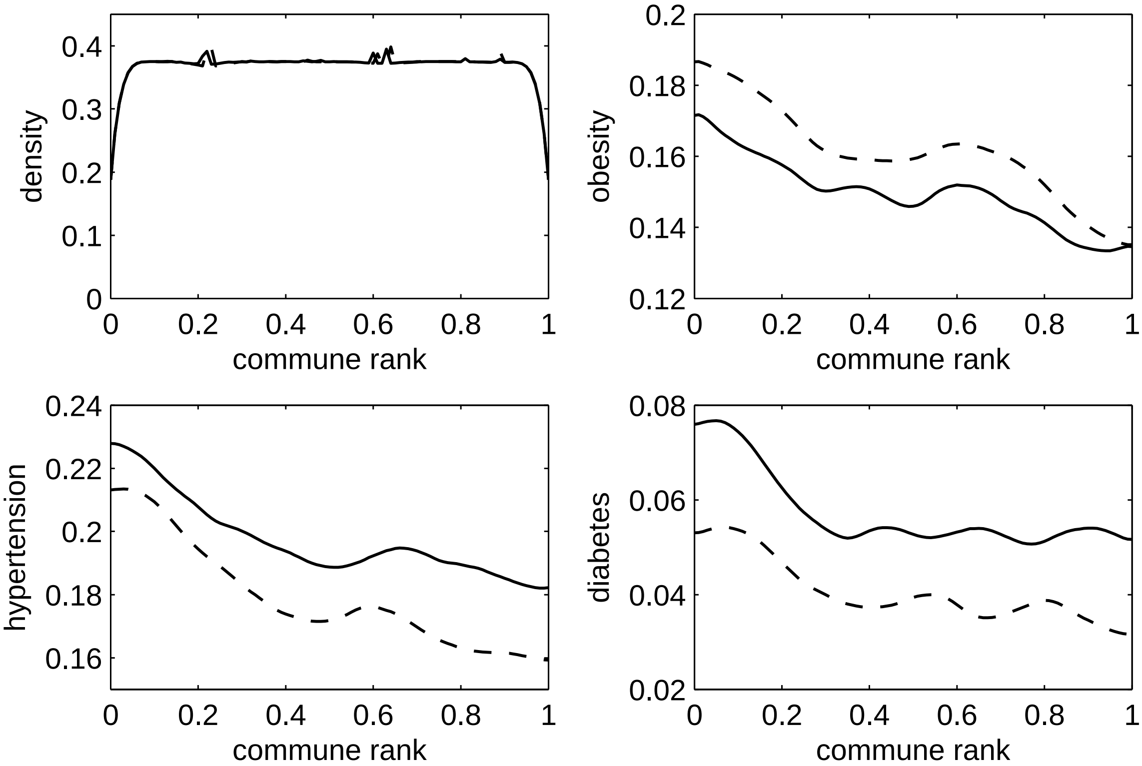


Figure 5.9: Deprivation-specific descriptive figure, deprivation index from factor analysis (pooled sample 2002, 2006)

Empirical density (upper left) and deprivation-specific (age-standardized) prevalences of obesity (upper right), hypertension (bottom left) and diabetes (bottom right) for the male (solid lines) and female (dashed lines) samples from the 2002 and 2006 Health Care Access Panel. Deprivation index derived from factor analysis.

analysis based) is the administrative area of Darmstadt. The best-off peak around 0.85 (0.89 factor analysis based) can be associated to the administrative area of Munich.

The results for obesity, hypertension and diabetes in figures 5.8 and 5.9 illustrate the results from tables 5.3 and 5.4. The age-standardized deprivation-specific estimates vary around their homogeneous counterparts shown in table 5.3. The patterns of the prevalence rates' deprivation-specific variations correspond to the significant health gradients to the detriment of deprived communes identified in table 5.4.[7] This is confirmed in the figures where higher community ranks (indicating less deprivation) are associated with lower prevalence rates. One may further note that the deprivation-specific variations of the prevalence rates are considerably lower than the age-specific variations demonstrated in figure 5.1. The comparably low variations are considered as negligible here and the prevalence correction applied to the unstandardized prevalences in previous chapters (see chapters 2, 3 and 4 as well as section 5.4.3 in this chapter) is omitted in this section. This seems justified as the correction factor $1/(1-\mu(z))$ from Wagstaff's correction formula (see equation 2.6 and section 5.3) would vary only marginally between different community ranks.

The distribution of mean net equivalent household income across deprivation ranks in figures 5.10 and 5.11 demonstrates that households in better-off communes have, on average, higher net equivalent incomes. This is consistent with the community deprivation-related concentration indices of income in table 5.4 and may also be explained by the comparably high weights assigned to per capita income in the theory-based and factor analysis based deprivation indices. Figures 5.10 and 5.11 also suggest that income inequality is somewhat higher in the better-off communes compared with the worse-off. This variation of the estimated deprivation-specific income inequalities is, however, only marginal and not statistically significant. The deprivation-specific estimates for both, mean net equivalent household income and income inequality, vary around their homogeneous counterparts shown in table 5.3.

[7] Note that the results for the theory-based weighting scheme index were also presented in table 5.3 but are identical to the left column of table 5.4.

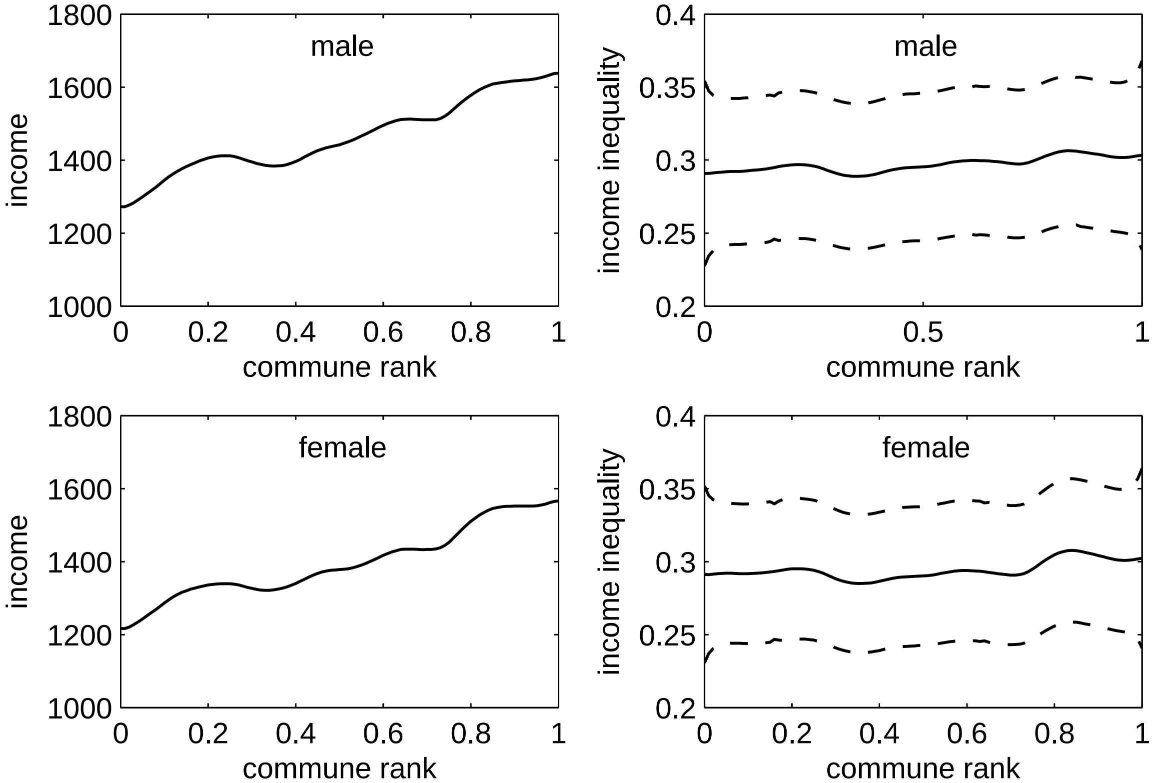


Figure 5.10: Deprivation-specific income and Gini indices, deprivation index from theory (pooled sample 2002, 2006)

Deprivation-specific income (left) and income inequality (right) for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel. Deprivation index derived from the literature.

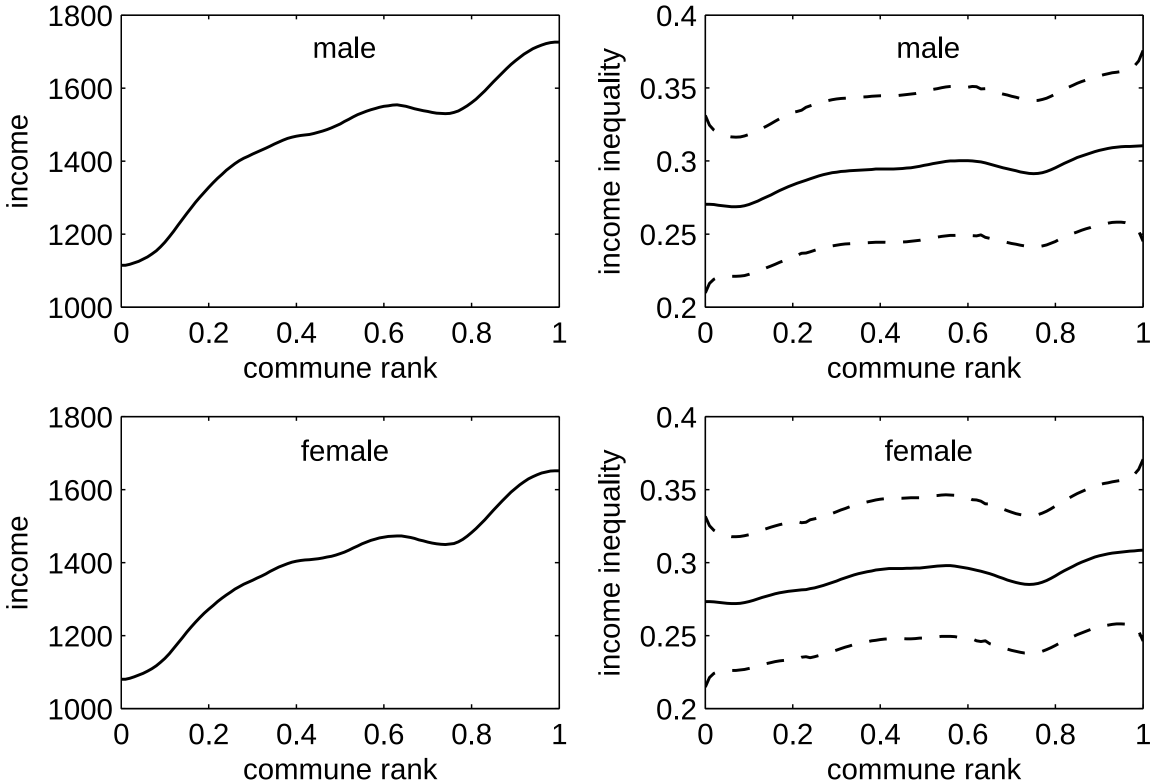


Figure 5.11: Deprivation-specific income and Gini indices, deprivation index from factor analysis (pooled sample 2002, 2006)

Deprivation-specific income (left) and income inequality (right) for the male (upper) and female (bottom) samples from the 2002 and 2006 Health Care Access Panel. Deprivation index derived from factor analysis.

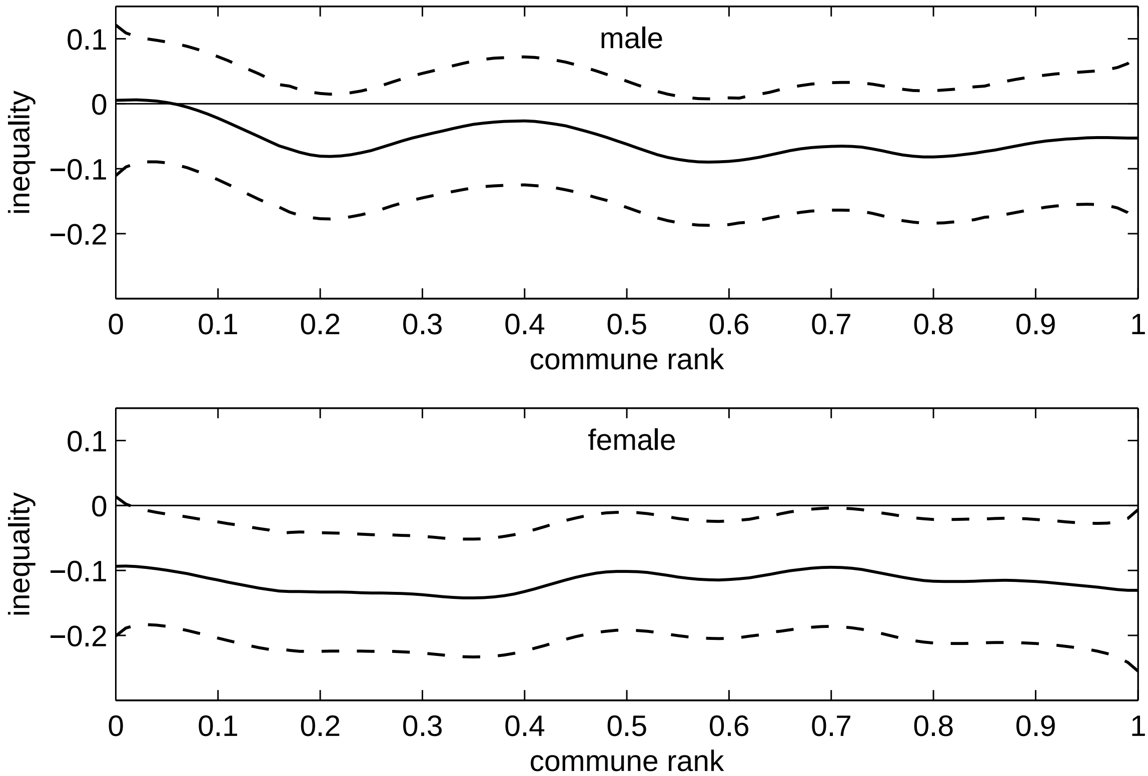


Figure 5.12: Deprivation-specific inequality of obesity, deprivation index from theory (pooled sample 2002, 2006)
Deprivation-specific income-related inequality (solid lines) of obesity with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples of the 2002 and 2006 Health Care Access Panel. Deprivation index derived from the literature.

Figures 5.12 and 5.13 illustrate the estimates for the deprivation-specific income-related inequalities in obesity. The estimated index for the male sample in figure 5.12 approximately equals zero for the worst-off five percent and varies around its homogeneous counterpart (-0.0592) from table 5.3. For the female sample, the index is significant for all deprivation ranks and, alike the index for the male sample, varies around its homogeneous counterpart in table 5.3 (i.e. -0.1209). The results for the factor analysis based weighting scheme are similar to those for the theory-based weighting scheme.

The results for deprivation-specific income-related inequalities in hypertension are shown in

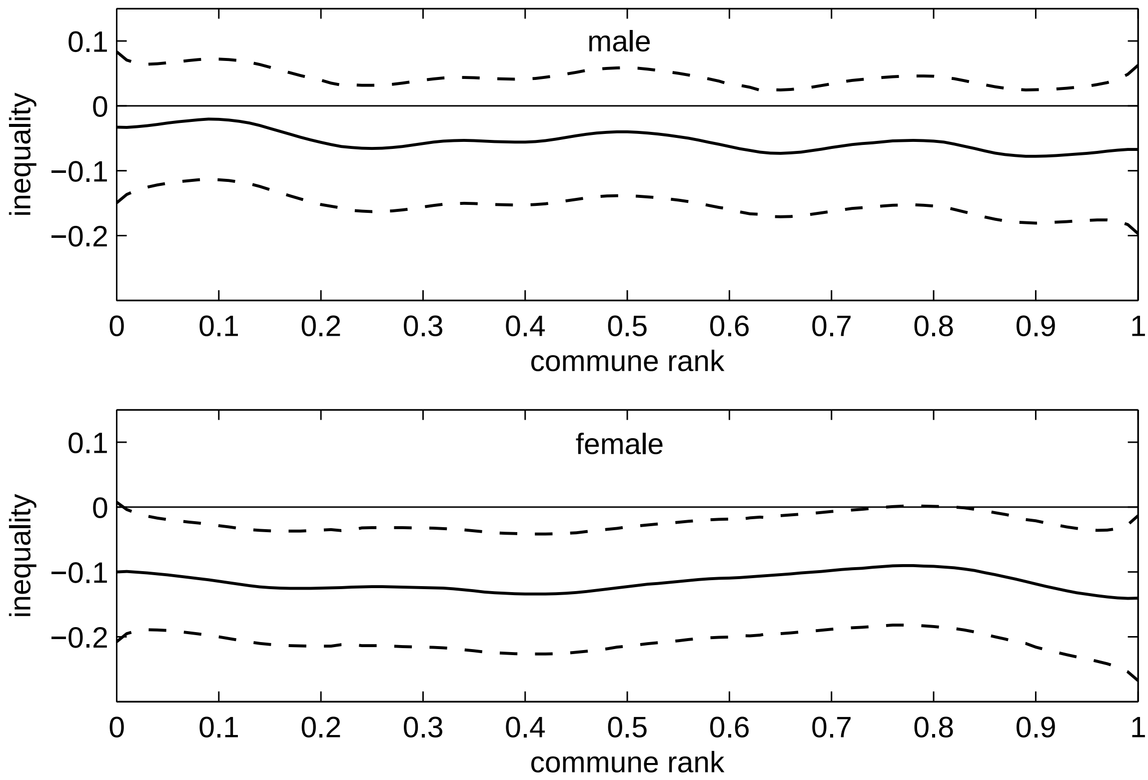


Figure 5.13: Deprivation-specific inequality of obesity, deprivation index from factor analysis (pooled sample 2002, 2006)
Deprivation-specific income-related inequality (solid lines) of obesity with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples of the 2002 and 2006 Health Care Access Panel. Deprivation index derived from factor analysis.

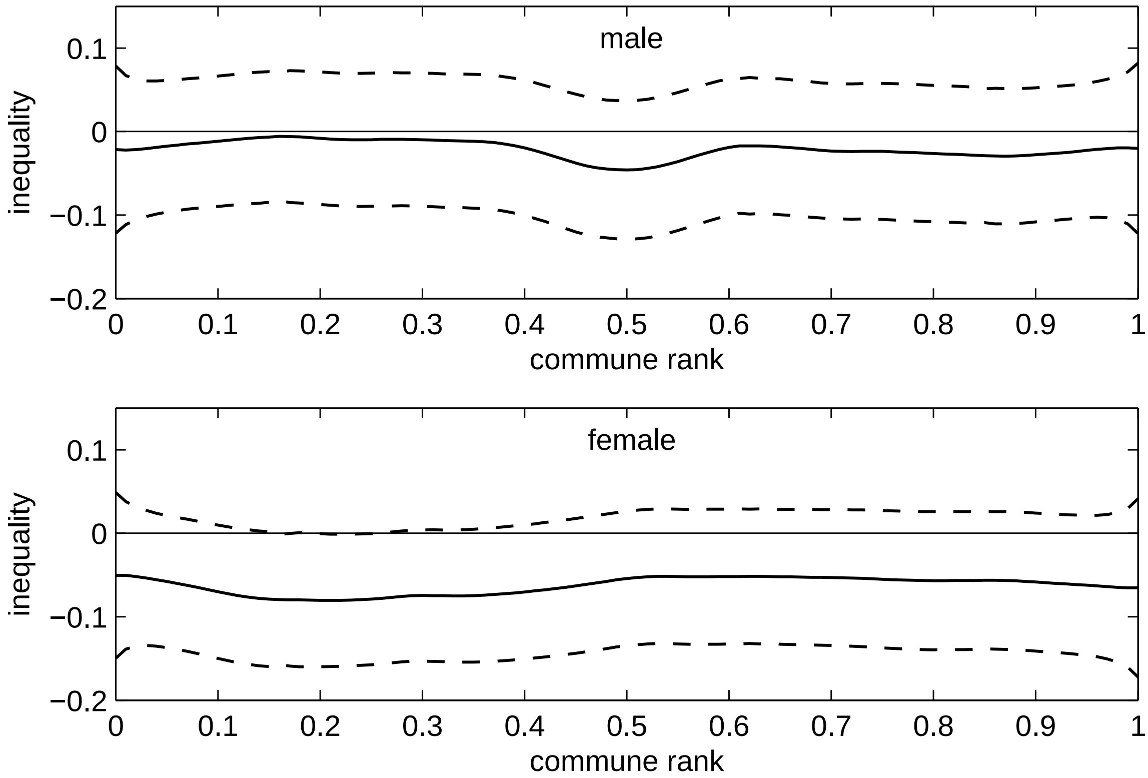


Figure 5.14: Deprivation-specific inequality of hypertension, deprivation index from theory (pooled sample 2002, 2006)
Deprivation-specific income-related inequality (solid lines) of hypertension with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples of the 2002 and 2006 Health Care Access Panel. Deprivation index derived from the literature.

figures 5.14 and 5.15. Similar to the results for obesity, the deprivation-specific estimates vary around their homogeneous counterparts. As noted before with respect the results for obesity, no considerable variations of the deprivation-specific inequalities are found for hypertension in both, the male and female samples.

The results for diabetes in figures 5.16 and 5.17 vary, similarly to the above described results for obesity and hypertension, around their homogeneous counterparts from table 5.3. The results suggest a slightly weaker inequality among the better-off 50 percent and particularly among the best-off communes. One may, however, agree that the decrease in inequality is

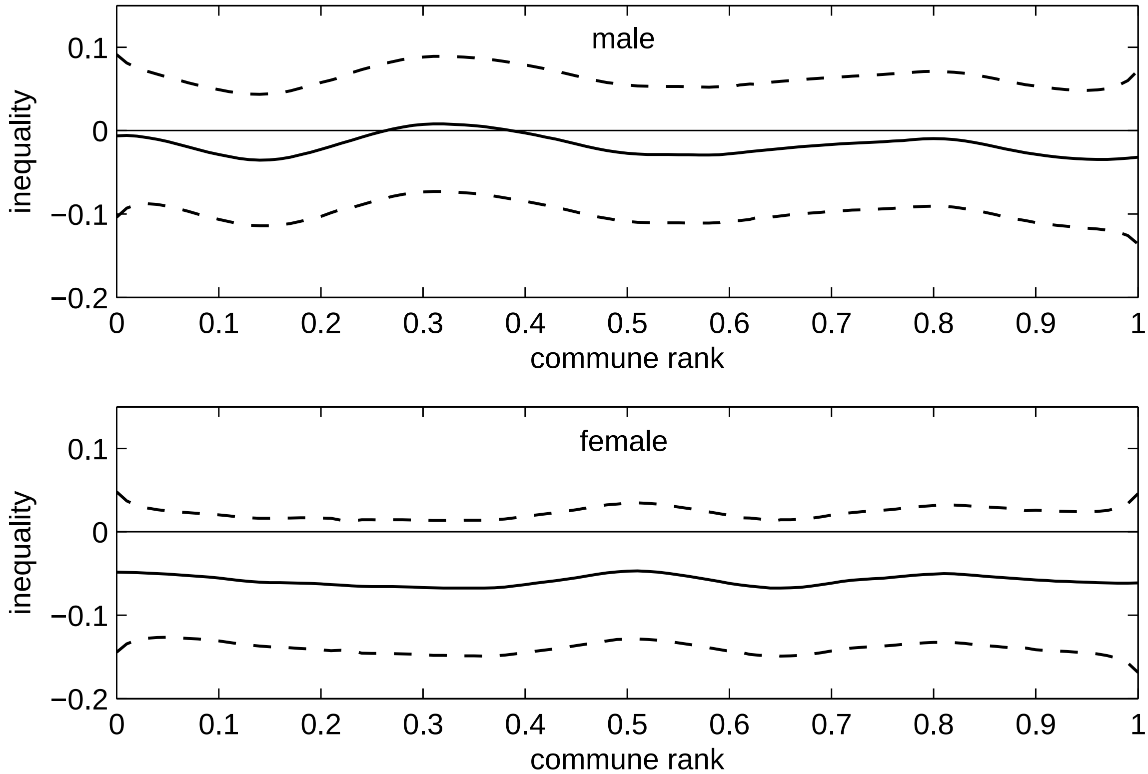


Figure 5.15: Deprivation-specific inequality of hypertension, deprivation index from factor analysis (pooled sample 2002, 2006)

Deprivation-specific income-related inequality (solid lines) of hypertension with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples of the 2002 and 2006 Health Care Access Panel. Deprivation index derived from factor analysis.

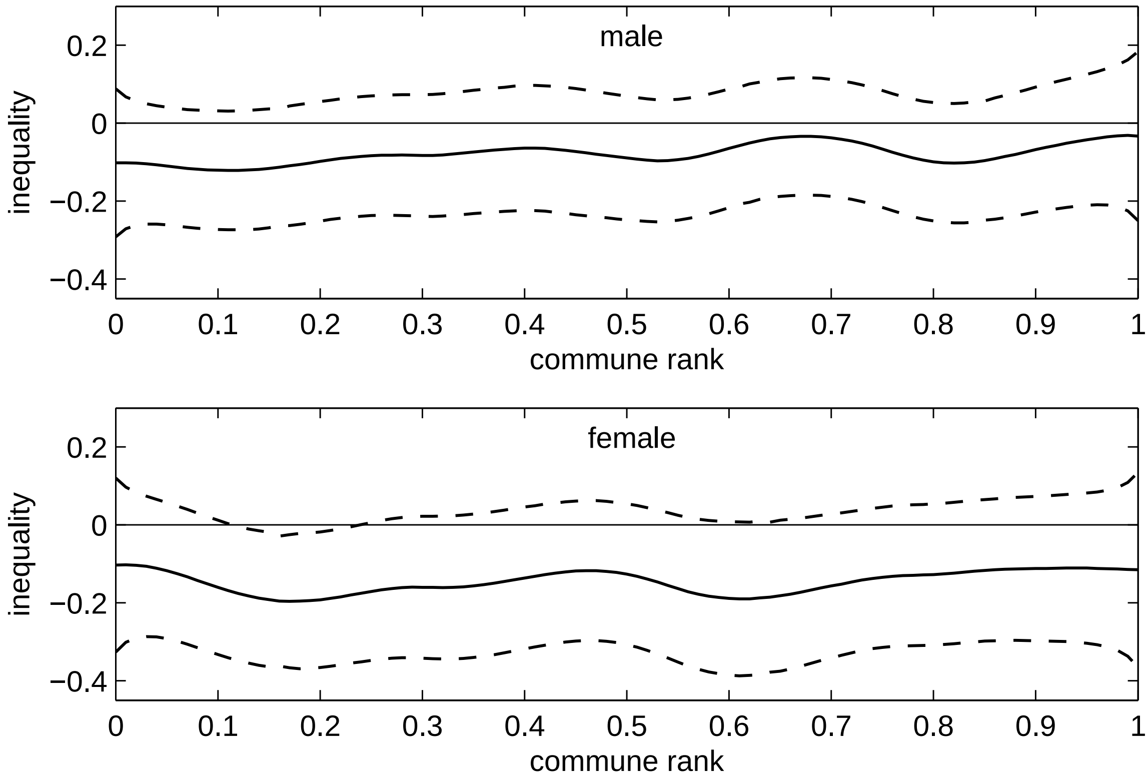


Figure 5.16: Deprivation-specific inequality of diabetes, deprivation index from theory (pooled sample 2002, 2006)
Deprivation-specific income-related inequality (solid lines) of diabetes with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples of the 2002 and 2006 Health Care Access Panel. Deprivation index derived from the literature.

only marginal and may not be considered as a significant result.

5.5 Discussion

The present chapter applied the Index of Multiple Deprivation which was adapted for Bavaria by Maier et al. (2011) and recently extended to a Germany-wide index. In addition to the theory-based weighting scheme applied by Maier et al. (2011), this chapter also applied a factor analysis based weighting scheme. Although the weights from these two approaches differ to some extent, the results are fairly similar. Both indices are highly correlated and

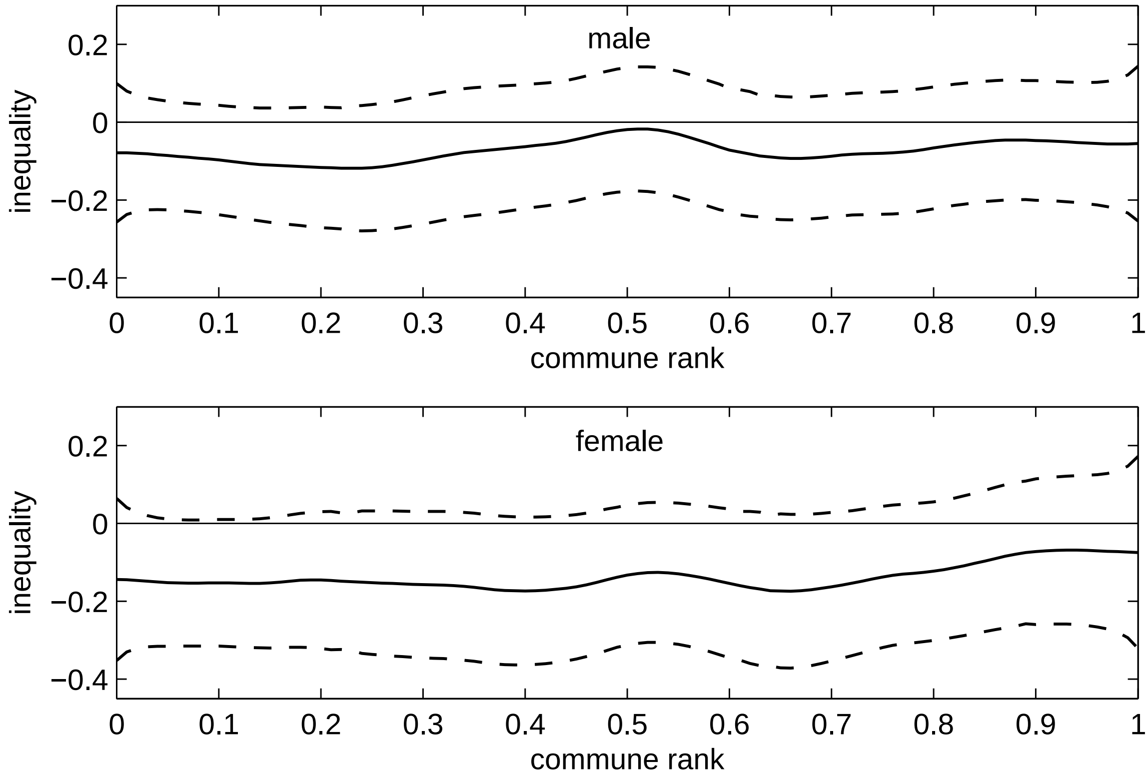


Figure 5.17: Deprivation-specific inequality of diabetes, deprivation index from factor analysis (pooled sample 2002, 2006)
Deprivation-specific income-related inequality (solid lines) of diabetes with 95 percent confidence intervals (dashed lines) for the male (upper) and female (bottom) samples of the 2002 and 2006 Health Care Access Panel. Deprivation index derived from factor analysis.

inequalities with respect to either of them have been found to differ only marginally. The rank sensitivity tests in table 5.4 suggest that changing the weighting scheme significantly changes the results for income for both sexes; as well as for hypertension and diabetes in the female sample.

Applying the varying inequality index introduced in chapter 2 has shown that the results for age-specific community deprivation-related inequalities are similar to those found when using net equivalent household income. The results for the community deprivation-specific variations of income-related health inequalities suggest no significant variations between deprivation ranks. The estimated inequalities are similar to their homogeneous counterparts indicating that the income-related health gradients do not vary with respect to relative community deprivation.

When analyzing the different weighting schemes in table 5.1, one may have noted that the theory-based weighting scheme proposed by Maier et al. (2011) and the weighting scheme obtained through factor analysis differ considerably in some points. What has been defined and considered as environmental deprivation by Maier et al. (2011) enters the Index of Multiple Deprivation as an advantage when applying the empirically driven weighting scheme. A possible explanation may be that the indicators assigned to the environmental deprivation domain may actually measure the degree of urbanization. While urbanized areas have environmental disadvantages compared with more rural areas, they may be considerably advantaged in other domains, particularly income and employment. A similar issue is observed concerning the education domain. Higher deprivation according to the theory-based weighting scheme is considered as an advantage by the factor analysis approach. Higher rates of unskilled workers may indicate a higher demand in the labor market and hence better opportunities particularly for unskilled persons in search of work. The education domain defined in Maier et al. (2011) may therefore alternatively be considered as an indicator of labor market advantages instead of being considered as an indicator of educational deprivation.

For the comparisons of income with the Indices of Multiple Deprivation (regardless of the underlying weighting scheme), one should note that communal income deprivation and net equivalent household income are not fully independent. Although the deprivation indices were not computed from the Health Care Access Panel, household incomes are still included in the indices as they are part of the average per tax payer income used as income domain (see table 5.1).

As already discussed in chapter 3 for self-reported data on specific diseases, one may argue that these would have to be diagnosed by a physician. As a consequence, one could expect reporting biases where distinct inequalities in health care utilization are observed. However, approximately 90 percent of the German population contact a physician within a year and Germany is known for its equitable access to health care (see e.g. van Doorslaer et al., 2004, 2006). Biases owing to inequalities in health care utilization hence seem rather unlikely. The potential of biases arising from social distances between physicians and less educated or lower income patients, however, remains. Considering the results found in Kelly-Irving et al. (2011), one may speculate that this would most likely lead to an underestimation of the concentration among the worse-off. Concerning the results for obesity, it should be mentioned that self-reported anthropometric data may involve some measurement or reporting errors which are again likely to lead to an underestimation of the prevalence of obesity.

When comparing socioeconomic gradients between age groups, using cross sectional data involves some further limitations. Some argue that variations in age-specific health inequalities, particularly leveling in older age groups, may be an artificial result owing to mortality selection (Dupre, 2007; Prus, 2004). It has been shown that this needs not necessarily be true for self reported health (Beckett, 2000), though. The age-specific mortality rates did not exceed 0.5 (1) percent for women younger than 60 (65) and were about double for men in the respective years in Germany (Human Mortality Database, 2012). The notion that the observed variations in income-related inequalities in older age groups could be solely caused by mortality selection hence seems rather unlikely (see also section 3.5).

Bibliography

Balia, S. and Jones, A. M. (2008). Mortality, lifestyle and socio-economic status. *Journal of Health Economics* **27**(1): 1–26.
URL: *http://dx.doi.org/10.1016/j.jhealeco.2007.03.001*

Banks, J., Marmot, M., Oldfield, Z. and Smith, J. P. (2007). The SES health gradient on both sides of the Atlantic. *IZA Discussion Papers 2539*, Institute for the Study of Labor (IZA).
URL: *http://ideas.repec.org/p/iza/izadps/dp2539.html*

Bauer, T., Göhlmann, S. and Sinning, M. (2007). Gender differences in smoking behavior. *Health Economics* **16**(9): 895–909.
URL: *http://dx.doi.org/10.1002/hec.1259*

Beckett, M. (2000). Converging health inequalities in later life–an artifact of mortality selection. *Journal of Health and Social Behavior* **41**(1): 106–119.
URL: *http://www.jstor.org/stable/pdfplus/2676363.pdf?acceptTC=true*

Bilger, M. (2008). Progressivity, horizontal inequality and reranking caused by health system financing: A decomposition analysis for Switzerland. *Journal of Health Economics* **27**(6): 1582–1593.
URL: *http://dx.doi.org/10.1016/j.jhealeco.2008.07.009*

Boersch-Supan, A. and Wilke, C. B. (2004). The German public pension system: How it was, how it will be. *Working Paper 10525*, National Bureau of Economic Research.

URL: *http://www.nber.org/papers/w10525*

Brockmann, H., Müller, R. and Helmert, U. (2009). Time to retire - time to die? A prospective cohort study of the effects of early retirement on long-term survival. *Social Science and Medicine* **69**(2): 160 – 164.

URL: *http://www.sciencedirect.com/science/article/pii/S0277953609002378*

Case, A. and Deaton, A. S. (2005). *Broken Down by Work and Sex: How Our Health Declines*. University of Chicago Press, pp. 185–212.

URL: *http://www.nber.org/chapters/c10361*

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

URL: *http://ideas.repec.org/a/aea/aecrev/v88y1998i2p248-53.html*

Dupre, M. E. (2007). Educational differences in age-related patterns of disease: Reconsidering the cumulative disadvantage and age-as-leveler hypotheses. *Journal of Health and Social Behavior* **48**(1): 1–15.

URL: *http://dx.doi.org/10.1177/002214650704800101*

Elo, I. T. and Preston, S. H. (1996). Educational differentials in mortality: United States, 1979-85. *Social Science and Medicine* **42**(1): 47–57.

URL: *http://www.sciencedirect.com/science/article/pii/0277953695000623*

Erreygers, G. (2009). Correcting the concentration index. *Journal of Health Economics* **28**(2): 504–515.

URL: *http://dx.doi.org/10.1016/j.jhealeco.2008.02.003*

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The*

Annals of Statistics **20**(4): pp. 2008–2036.
URL: *http://www.jstor.org/stable/2242378*

FSO (2006). Lebenslagen in Deutschland. *Federal Statistical Office*.
URL: *http:/ / www.destatis.de/ jetspeed/ portal/ cms/ Sites/ destatis/ Internet/ DE/ Presse/ pk/ 2006/ Mikrozensus/ Pressebroschuere,property=file.pdf*

Genuneit, J., Weinmayr, G., Radon, K., Dressel, H., Windstetter, D., Rzehak, P., Vogelberg, C., Leupold, W., Nowak, D., von Mutius, E. and Weiland, S. K. (2006). Smoking and the incidence of asthma during adolescence: Results of a large cohort study in Germany. *Thorax* **61**(7): 572–578.
URL: *http://dx.doi.org/10.1136/thx.2005.051227*

Giskes, K., Kunst, A. E., Benach, J., Borrell, C., Costa, G., Dahl, E., Dalstra, J. A. A., Federico, B., Helmert, U., Judge, K., Lahelma, E., Moussa, K., Ostergren, P. O., Platt, S., Prattala, R., Rasmussen, N. K. and Mackenbach, J. P. (2005). Trends in smoking behaviour between 1985 and 2000 in nine European countries by education. *Journal of Epidemiology and Community Health* **59**(5): 395–401.
URL: *http://dx.doi.org/10.1136/jech.2004.025684*

Graham, H. (1996). Smoking prevalence among women in the European community 1950-1990. *Social Science and Medicine* **43**(2): 243–254.
URL: *http://www.sciencedirect.com/science/article/pii/027795369500369X*

Haffner, S. M. (2006). Relationship of metabolic risk factors and development of cardiovascular disease and diabetes. *Obesity (Silver Spring)* **14 Suppl 3**: 121S–127S.
URL: *http://dx.doi.org/10.1038/oby.2006.291*

Harati, H., Hadaegh, F., Momenan, A. A., Ghanei, L., Bozorgmanesh, M. R., Ghanbarian, A., Mirmiran, P. and Azizi, F. (2010). Reduction in incidence of type 2 diabetes by lifestyle

intervention in a Middle Eastern community. *American Journal of Preventive Medicine* **38**(6): 628–636.
URL: *http://dx.doi.org/10.1016/j.amepre.2010.03.003*

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)* **55**(4): 757–796.
URL: *http://www.jstor.org/stable/2345993*

Helmert, U. and Buitkamp, M. (2004). Changes in smoking habits in Germany between 1985 and 2002. *Gesundheitswesen* **66**: 102–6.
URL: *http://www.thieme-connect.com/ejournals/pdf/gesu/doi/10.1055/s-2004-812824.pdf*

Human Mortality Database (2011). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at `www.mortality.org` or `www.humanmortality.de` (data downloaded on June 20, 2012).

Humphries, K. H. and van Doorslaer, E. (2000). Income-related health inequality in Canada. *Social Science and Medicine* **50**(5): 663–671.
URL: *http://www.sciencedirect.com/science/article/pii/S0277953699003196*

Islam, M. K., Gerdtham, U.-G., Clarke, P. and Burström, K. (2010). Does income-related health inequality change as the population ages? Evidence from Swedish panel data. *Health Economics* **19**(3): 334–349.
URL: *http://dx.doi.org/10.1002/hec.1479*

Jones, A. and López Nicolás, A. (2006). Allowing for heterogeneity in the decomposition of measures of inequality in health. *Journal of Economic Inequality* **4**(3): 347–365.
URL: *http://ideas.repec.org/a/kap/jecinq/v4y2006i3p347-365.html*

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

103.
URL: *http://ideas.repec.org/a/eee/econom/v77y1997i1p87-103.html*

Kamholz, S. L. (2004). Pulmonary and cardiovascular consequences of smoking. *Medical Clinics of North America* **88**(6): 1415 – 1430.
URL: *http://www.sciencedirect.com/science/article/B75JC-4DFJN7H-6/2/151f7c2976a88bbf20acb175e3f901e3*

Kelly-Irving, M., Delpierre, C., Schieber, A.-C., Lepage, B., Rolland, C., Afrité, A., Pascal, J., Cases, C., Lombrail, P. and Lang, T. (2011). Do general practitioners overestimate the health of their patients with lower education? *Social Science and Medicine* **73**(9): 1416–1421.
URL: *http://dx.doi.org/10.1016/j.socscimed.2011.07.031*

Kim, J. and Durden, E. (2007). Socioeconomic status and age trajectories of health. *Social Science and Medicine* **65**(12): 2489–2502.
URL: *http://dx.doi.org/10.1016/j.socscimed.2007.07.022*

Kim, J. and Miech, R. (2009). The Black-White difference in age trajectories of functional health over the life course. *Social Science and Medicine* **68**(4): 717–725.
URL: *http://dx.doi.org/10.1016/j.socscimed.2008.12.021*

Konings, P., Harper, S., Lynch, J., Hosseinpoor, A. R., Berkvens, D., Lorant, V., Geckova, A. and Speybroeck, N. (2010). Analysis of socioeconomic health inequalities using the concentration index. *International Journal of Public Health* **55**(1): 71–74.
URL: *http://dx.doi.org/10.1007/s00038-009-0078-y*

Kunst, A. E. and Mackenbach, J. P. (1994). The size of mortality differences associated with educational level in nine industrialized countries. *American Journal of Public Health*

84(6): 932–937.
URL: *http://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.84.6.932*

Kuznetsov, L., Maier, W., Hunger, M., Meyer, M. and Mielck, A. (2011). Associations between regional socioeconomic deprivation and cancer risk: Analysis of population-based cancer registry data from Bavaria, Germany. *Preventive Medicine* **53**(4-5): 328–330.
URL: *http://dx.doi.org/10.1016/j.ypmed.2011.07.021*

Kuznetsov, L., Maier, W., Hunger, M., Meyer, M. and Mielck, A. (2012). Regional deprivation in Bavaria, Germany: Linking a new deprivation score with registry data for lung and colorectal cancer. *International Journal of Public Health* **e-pub ahead of print**.
URL: *http://dx.doi.org/10.1007/s00038-012-0342-4*

Lampert, T. (2010). Social determinants of tobacco consumption among adults in Germany. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **53**(2): 108–116.
URL: *http://dx.doi.org/10.1007/s00103-009-1014-2*

Lampert, T. and Burger, M. (2004). Smoking habits in Germany – results of the German National Telephone Health Survey 2003. *Gesundheitswesen* **66**: 511–517.
URL: *http://dx.doi.org/10.1055/s-2004-813527*

Lampert, T. and Thamm, M. (2004). Social inequality and smoking behavior in Germany. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **47**(11): 1033–1042.
URL: *http://dx.doi.org/10.1007/s00103-004-0934-0*

Lechert, Y. and Schimpl-Neimanns, B. (2007). GESIS-ZUMA-Methodenbericht 2007/08 Mikrozensus Scientific Use File 2005 Dokumentation und Datenaufbereitung. *Federal Statistical Office*.
URL: *http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2007/07_08_lechert.pdf*

Lerman, R. I. and Yitzhaki, S. (1989). Improving the accuracy of estimates of Gini coefficients. *Journal of Econometrics* **42**(1): 43–47.
URL: *http://ideas.repec.org/a/eee/econom/v42y1989i1p43-47.html*

Li, Q., Huang, C. J., Li, D. and Fu, T.-T. (2002). Semiparametric smooth coefficient models. *Journal of Business and Economic Statistics* **20**(3): 412–22.
URL: *htp://www.jstor.org/stable/1392126*

Lindelow, M. (2006). Sometimes more equal than others: How health inequalities depend on the choice of welfare indicator. *Health Economics* **15**(3): 263–279.
URL: *http://dx.doi.org/10.1002/hec.1058*

Lüngen, M., Siegel, M., Drabik, A., Büscher, G. and von Törne, I. (2009). Ausmaß und Gründe für Ungleichheiten der gesundheitlichen Versorgung in Deutschland. *Studien zu Gesundheit, Medizin und Gesellschaft* **05/2009**.
URL: *http://gesundheitsoekonomie.uk-koeln.de/forschung/schriftenreihe-sgmg/sgmg*

Lynch, S. (2003). Cohort and life-course patterns in the relationship between education and health: A hierarchical approach. *Demography* **40**: 309–331. 10.1353/dem.2003.0016.
URL: *http://dx.doi.org/10.1353/dem.2003.0016*

Maier, W., Fairburn, J. and Mielck, A. (2011). Regional Deprivation and mortality in Bavaria. Development of a community-based Index of Multiple Deprivation. *Gesundheitswesen* **e-pub ahead of print**.
URL: *https://www.thieme-connect.com/DOI/DOI?10.1055/s-0031-1280846*

McKinnon, B., Harper, S. and Moore, S. (2011). Decomposing income-related inequality in cervical screening in 67 countries. *International Journal of Public Health* **56**(2): 139–152.
URL: *http://dx.doi.org/10.1007/s00038-010-0224-6*

Mirowsky, J. and Ross, C. (2005). Education, cumulative advantage, and health. *Ageing International* **30**: 27–62.
URL: *http://dx.doi.org/10.1007/BF02681006*

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**(3): 703–708.
URL: *http://www.jstor.org/stable/1913610*

Noble, M., Wright, G., Smith, G. and Dibben, C. (2006). Measuring multiple deprivation at the small-area level. *Environment and Planning* **38**(1): 169–185.
URL: *http://dx.doi.org/10.1068/a37168*

O'Donnell, O., van Doorslaer, E., Wagstaff, A. and Lindelow, M. (2008). *Analyzing Health Equity using Household Survey Data*. WBI Learning Resources Series, World Bank Institute.

Peto, R., Darby, S., Deo, H., Silcocks, P., Whitley, E. and Doll, R. (2000). Smoking, smoking cessation, and lung cancer in the UK since 1950: Combination of national statistics with two case-control studies. *BMJ: British Medical Journal* **321**(7257): 323–329.
URL: *http://dx.doi.org/10.1136/bmj.321.7257.323*

Potthoff, P., Heinemann, L. A. J. and Güther, B. (2004). A household panel as a tool for cost-effective health-related population surveys: Validity of the "Healthcare Access Panel". *German Medical Science* **2**: 1–8.
URL: *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2703211/pdf/GMS-02-05.pdf*

Prus, S. G. (2004). A life course perspective on the relationship between socio-economic status and health: Testing the divergence hypothesis. *Canadian Journal on Aging* **23 Suppl 1**: S145–S153.
URL: *http://dx.doi.org/10.1353/cja.2005.0040*

Puska, P. (2010). From Framingham to North Karelia: From descriptive epidemiology to public health action. *Progress in Cardiovascular Diseases* **53**(1): 15–20.
URL: *http://dx.doi.org/10.1016/j.pcad.2010.01.003*

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

Reeske, A., Spallek, J. and 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.
URL: *http://dx.doi.org/10.1186/1475-9276-8-26*

Richter, M. and Leppin, A. (2007). Trends in socio-economic differences in tobacco smoking among German schoolchildren, 1994-2002. *European Journal of Public Health* **17**(6): 565–571.
URL: *http://dx.doi.org/10.1093/eurpub/ckm010*

Ross, C. E. and Wu, C. L. (1996). Education, age, and the cumulative advantage in health. *Journal of Health and Social Behavior* **37**(1): 104–120.
URL: *http://www.jstor.org/stable/2137234*

Schaap, M. M. and Kunst, A. E. (2009). Monitoring of socio-economic inequalities in smoking: Learning from the experiences of recent scientific studies. *Public Health* **123**(2): 103–109.
URL: *http://dx.doi.org/10.1016/j.puhe.2008.10.015*

Schulze, A. and Mons, U. (2006). The evolution of educational inequalities in smoking: A changing relationship and a cross-over effect among German birth cohorts of 1921-70. *Addiction* **101**(7): 1051–1056.
URL: *http://dx.doi.org/10.1111/j.1360-0443.2006.01458.x*

Sen, A. (2002). Why health equity? *Health Economics* **11**(8): 659–666.
URL: *http://dx.doi.org/10.1002/hec.762*

Slama, K. (2008). Global perspective on tobacco control. Part I. The global state of the tobacco epidemic [State Of The Art Series. Chronic obstructive pulmonary disease in high- and low-income countries. Edited by G. Marks and M. Chan-Yeung. Number 1 in the series]. *The International Journal of Tuberculosis and Lung Disease* **12**(1): 3–7.
URL: *http://www.ingentaconnect.com/content/iuatld/ijtld/2008/00000012/00000001/art00003*

Sundmacher, L. (2011). The effect of health shocks on smoking and obesity. *European Journal of Health Economics* **13**(4):451–460.
URL: *http://dx.doi.org/10.1007/s10198-011-0316-0*

van Doorslaer, E. and Koolman, X. (2004). Explaining the differences in income-related health inequalities across European countries. *Health Economics* **13**(7): 609–628.
URL: *http://dx.doi.org/10.1002/hec.918*

van Doorslaer, E., Koolman, X. and Jones, A. M. (2004). Explaining income-related inequalities in doctor utilisation in Europe. *Health Economics* **13**(7): 629–647.
URL: *http://dx.doi.org/10.1002/hec.919*

van Doorslaer, E., Masseria, C., Koolman, X. for the OECD Health Equity Research Group (2006). Inequalities in access to medical care by income in developed countries. *Canadian Medical Association Journal* **174**(2): 177–183.
URL: *http://www.cmaj.ca/content/174/2/177.abstract*

van Doorslaer, E., Wagstaff, A., Bleichrodt, H., Calonge, S., Gerdtham, U.-G., Gerfin, M., Geurts, J., Gross, L., Häkkinen, U., Leu, R. E., O'Donell, O., Propper, C., Puffer, F., Rodríguez, M., Sundberg, G. and Winkelhake, O. (1997). Income-related inequalities in

health: Some international comparisons. *Journal of Health Economics* **16**(1): 93 – 112.
URL: *http://www.sciencedirect.com/science/article/pii/S0167629696005322*

van Doorslaer, E., Wagstaff, A., van der Burg, H., Christiansen, T., Graeve, D. D., Duchesne, I., Gerdtham, U.-G., Gerfin, M., Geurts, J., Gross, L., Häkkinen, U., John, J., Klavus, J., Leu, R. E., Nolan, B., O'Donnell, O., Propper, C., Puffer, F., Schellhorn, M., Sundberg, G. and Winkelhake, O. (2000). Equity in the delivery of health care in Europe and the US. *Journal of Health Economics* **19**(5): 553 – 583.
URL: *http://www.sciencedirect.com/science/article/pii/S0167629600000503*

van Kippersluis, H., O'Donnell, O., van Doorslaer, E. and van Ourti, T. (2010). Socioeconomic differences in health over the life cycle in an egalitarian country. *Social Science and Medicine* **70**(3): 428–438.
URL: *http://dx.doi.org/10.1016/j.socscimed.2009.10.020*

van Kippersluis, H., van Ourti, T., O'Donnell, O. and van Doorslaer, E. (2009). Health and income across the life cycle and generations in Europe. *Journal of Health Economics* **28**(4): 818–830.
URL: *http://dx.doi.org/10.1016/j.jhealeco.2009.04.001*

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.
URL: *http://www.sciencedirect.com/science/article/B6V64-3VW8NC3-11/2/a7d85cb0dd16d438cd43d7c74e0e729f*

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): 429–432.
URL: *http://dx.doi.org/10.1002/hec.953*

Wagstaff, A. (2011). The concentration index of a binary outcome revisited. *Health Economics* **20**(10): 1155–1160.
URL: *http://dx.doi.org/10.1002/hec.1752*

Wagstaff, A., Paci, P. and van Doorslaer, E. (1991). On the measurement of inequalities in health. *Social Science and Medicine* **33**(5): 545 – 557.
URL: *http://www.sciencedirect.com/science/article/pii/027795369190212U*

Wagstaff, A. and van Doorslaer, E. (2000). Measuring and testing for inequity in the delivery of health care. *The Journal of Human Resources* **35**(4): 716–733.
URL: *http://www.jstor.org/stable/146369*

Wagstaff, A., van Doorslaer, E. and 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.
URL: *http://www.sciencedirect.com/science/article/B6VC0-46RKR42-2/2/779b6d6fc7e0ba130b2d346c0274d63f*

Wagstaff, A. and Watanabe, N. (2003). What difference does the choice of SES make in health inequality measurement? *Health Economics* **12**(10): 885–890.
URL: *http://dx.doi.org/10.1002/hec.805*

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**(4): 817–838.
URL: *http://www.jstor.org/stable/1912934*

WHO (2003). Obesity and overweight: Global strategy on diet, physical activity and health. *World Health Organization Fact Sheet* .
URL: *www.who.int/dietphysicalactivity/media/en/gsfs_obesity.pdf*

WHO (2009). Global Database on body mass index: BMI classification. *The World Health Organization* .
URL: *http://apps.who.int/bmi/index.jsp?introPage=intro_3.html, accessed September 15th, 2009*

Wildman, J. (2003). Income related inequalities in mental health in Great Britain: Analysing the causes of health inequality over time. *Journal of Health Economics* **22**(2): 295–312.
URL: *http://dx.doi.org/10.1016/S0167-6296(02)00101-7*

Willson, A. E., Shuey, K. M. and Elder, Jr., G. H. (2007). Cumulative advantage processes as mechanisms of inequality in life course health. *American Journal Of Sociology* **112**(6): 1886–1924.
URL: *http://www.journals.uchicago.edu/doi/abs/10.1086/512712*

Wingerter, C. (2010). Später in den Ruhestand? *Technical report*, German Federal Statistical Office, Wiesbaden.
URL: *https://www.destatis.de/DE/Publikationen/STATmagazin/Arbeitsmarkt/2010_09/2010_09PDF.pdf?__blob=publicationFile*

Appendix A

Curriculum Vitae

Martin Siegel

Persönliche Daten

Name:	Martin Siegel
Geboren am:	22. April 1980 in Gießen
Anschrift:	Friedrich-Ebert-Straße 30, 50354 Hürth

Schule und Studium

1986 – 1990	Grundschule Altwarmbüchen
1990 – 1992	Orientierungsstufe (5. und 6. Klasse) der Grundschule Altwarmbüchen
1992 – 1993	Gymnasiale Mittelstufe, Ricarda-Huch-Schule, Hannover
1993 – 1996	Gymnasiale Mittelstufe, Carl-Strehl-Schule, Marburg
1996 – 1999	Gymnasiale Oberstufe (*berufliches Gymnasium*), Carl-Strehl-Schule, Marburg
06/1999	Abitur (Allgemeine Hochschulreife, Fachrichtung Wirtschaft), Carl-Strehl-Schule, Marburg
09/1999 – 02/2000	Reedereilogistik an der Fachhochschule Ostfriesland in Leer
04/2000 – 12/2005	Diplomstudiengang Volkswirtschaftslehre mit Schwerpunkt quantitative Wirtschaftsforschung an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Christian-Albrechts-Universität zu Kiel

12/2005	Abschluss als Diplom-Volkswirt, Schwerpunkt quantitative Wirtschaftsforschung; Diplomarbeit im Fach Ökonometrie
10/2006 – 08/2007	Sprachkurse am Institut für Orientalistik der Christian-Albrechts-Universität zu Kiel
seit 04/2008	Promotionsstudium an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Albertus-Magnus-Universität zu Köln

Praktische Erfahrungen und sonstige Tätigkeiten

01/2006 – 04/2006	Praktikum bei *Jan Kala Sahitya Manch Sanstha* (JKSMS), einer Nichtregierungsorganisation (*NGO*) in Jaipur, Indien; Mitarbeit im *Endogenous Tourism Project*, einer Kooperation des United Nations Development Programme (*UNDP*), der indischen Regierung und JKSMS.
09/2006–12/2006	Freiwillige Mitarbeit am Lehrstuhl für Ökonometrie der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Christian-Albrechts-Universität zu Kiel: Überarbeitung der Diplomarbeit zwecks Publikation in englischer Sprache
seit 09/2007	Wissenschaftlicher Mitarbeiter am Institut für Gesundheitsökonomie und klinische Epidemiologie der Universität zu Köln