MEASURING SOCIO-ECONOMIC INEQUALITY IN ILL-HEALTH USING PERMANENT INCOME[#]

Tom Van Ourti[*]

UFSIA, University of Antwerp, Belgium

Abstract

In Belgium, income-related inequality in ill-health seems to favour the rich, meaning that the rich are generally in better health than the poor are. Restricting the analysis to subsamples of the Belgian population, slightly modifies the conclusion, i.e. there is no income-related inequality in ill-health among the 65+. Since it is not clear whether the absence in inequality stems from the limited variation in the income of the 65+ (because of welfare benefits) or whether it truly reflects reality, I did the analysis over again using estimates of permanent income instead of income. It turned out that inequality among the 65+ remained very limited indeed, yet robustness checks pointed to the fragility of the results.

JEL classification: I19, D31

Key words: inequality, ill-health, permanent income

I would like to thank D. De Graeve, G. Erreygers and participants in UFSIA-workshops for helpful hints.

* Corresponding author: Tom Van Ourti, Department of economics, UFSIA, University of Antwerp, Prinsstraat 13, B-2000 Antwerpen (Belgium). Tel: +32/3/220.41.40. Fax: +32/3/220.47.99. E-mail: tom.vanourti@ufsia.ac.be

1. Introduction

In this paper, I focus on socio-economic inequality in ill-health[1]. The research question boils down to the following: are the poorer members in better (worse) health than the richer members of society are? This question is of considerable importance. Indeed, the answer can be used to evaluate the performance of health care systems. More specifically, it can be used to evaluate the output of health care systems in terms of its capacity to reduce income-related inequity in ill-health. However, the merits should not be exaggerated. There is no focus on efficiency of health care systems, nor on the mean level of ill-health in a society. Moreover, a high level of inequity does not necessarily point to a bad performance of a health care system. Two distinct societies might have a different degree of inequity, but both health care systems might be equally performing in terms of reducing income-related inequity in ill-health. Differences in income inequality and income-related health behaviour (e.g. smoking) might account for the above. Notwithstanding the limitations of the research question, it remains interesting from a positive point of view. Knowing whether the poor are less (more) healthy than the rich, is interesting in itself and could help policy makers in formulating health care policy.

This analysis builds on two strands of literature: on the one hand, the permanent income hypothesis and on the other hand the existing literature on socio-economic inequality in ill-health.

The latter resulted for the most form the ECuity project. The ECuity project was founded in the late eighties and intends to look deeper into equity issues of health care

[1] Note that I use inequality in ill-health instead of inequality in health. In section 2.1, I describe the importance to distinguish.

systems within European countries (+US). The major aim is to do comparative analysis and to find causal linkages between characteristics and outcomes of health care systems. A specific methodology was developed to measure socio-economic inequality in ill-health. First, income, as a measure of socio-economic status, is used to stratify the sample. Next, concentration curves and indices are applied to the stratified sample to quantify the possibly existing inequity in ill-health. In all participating countries, a negative relationship was found between socio-economic status and ill-health status, i.e. the richer are in general in better health than the poorer are. (e.g., van Doorslaer et al., 1997; Humphries and van Doorslaer, 2000; De Graeve and Duchesne, 1997)

In this paper, the Belgian situation is reanalysed using a different methodology. Instead of using income as a measure of socio-economic status, I use permanent income. The use of an alternative measure was inspired by previous, but unpublished research[2]; i.e. there seems to be no *income*-related inequality in ill-health among persons of at least 65 years (65+). In contrast, significant inequity is found for the entire population. Two possible routes could be explored. First, one could do nothing and conclude that indeed there does not exist any income-related inequality in ill-health among the 65+. Second, one could try another proxy for socio-economic status and add the results to the existing evidence. Such a procedure might further clarify the relationship between socio-economic status and inequality in ill-health. In this analysis, I opted for permanent income. This choice was dictated by the limited variation in the income of the 65+ (because of welfare benefits), which might account for the absence of any inequality in ill-health. Using permanent income instead might

[2] Previous as well as current results are presented in this paper.

give a better idea of one's socio-economic status, certainly for the older members of society. Off course, one could opt for other proxies of socio-economic status as well. The obvious choice would be to complement the income data with data on income from financial assets, etc. Another possibility consists of computing concentration curves and indices for different characteristics of one's socio-economic status (e.g. education, income, etc). Nevertheless, given the data limitations, I opt for permanent incomes. They will be estimated using the methodology proposed by King & Dicks-Mireaux (1982).

The second section of this paper deals with the theoretical aspects. Concentration curves, indices and the estimating strategy to compute permanent incomes will be briefly discussed. The data are dealt with in the third section. Additional econometric theory and the estimation of permanent income in terms of observables are given in section four. The fifth part discusses the results and finally I conclude.

2. The methodologies applied

In this part, both methodologies are explained. In the first section, I describe the ECuity methodology, which is applied except for the stratification variable, namely income. Instead, I use permanent income. The estimating strategy is dealt with in the second section.

2.1. Measuring socio-economic inequality in ill-health

As explained in the introduction, a specific methodology to quantify socio-economic inequality in ill-health was developed for the ECuity project. Wagstaff et al. (1991) compared different inequality measures frequently used in the income inequality literature. They concluded that concentration curves and indices are best suited to measure socio-economic inequality in ill-health. First of all, concentration curves and indices reflect inequality in ill-health and emphasise its relation with the socio-economic dimension. Second, they take the experience – in terms of ill-health and income – of the entire population into account. Third, changes in the distribution of socio-economic status (e.g. a transfer of ill-health from a disadvantaged to an advantaged person) are reflected. Fourth, concentration curves and indices can be visually represented. Finally, both measures "have a firm grounding in the literature on income distribution and redistribution – its properties and the value judgements underlying it are therefore reasonably well understood". (cit., van Doorslaer et al., 1997)

In the following, I briefly discuss the concentration curve and index. In a first step, one must stratify the sample by socio-economic status. This implies that we rank the individuals by socio-economic status, beginning with the least advantaged. Next, supposing we have a continuous measure of ill-health, we can construct an illness concentration curve [L(p)]. L(p) plots the cumulative proportion of ill-health against the cumulative proportion of the population, ranked by socio-economic status. From the definition it is immediately clear that we face a relative measure, i.e. L(p) is independent of the mean level of income and ill-health, and proportional changes in ill-health do not change the concentration curve. In figure 1, an illness concentration curve is depicted.

[Figure 1]

If the concentration curve lies above the diagonal, the more disadvantaged members of society have a higher share of total ill-health. In other words, they are less healthy than the advantaged members of society are. (pro-rich inequity) If the concentration curve lies entirely underneath the diagonal, the opposite can be concluded. (pro-poor inequity) Coincidence of the concentration curve and the diagonal reflects socio-economic equality in ill-health. A last possibility concerns an intersection between the concentration curve and the diagonal. In this case, compensation arises between pro-rich and pro-poor inequity.

One can compare inequality in different (sub) populations by means of two or more concentration curves. If a concentration curve is dominated, one could conclude that there is unambiguously less inequality in ill-health.[3] The inequality ranking is, however, partial in the sense that one cannot rank intersecting curves. In that case, one might calculate a concentration index (CI) that yields a complete ordering. CI is defined as twice the area between L(p) and the diagonal. Since I use grouped data (see section 5), we have the following definition (assuming that L(p) is piecewise linear):

[3] "We conjecture that it would be possible to prove an analogue of the theorem of Atkinson for the case of health inequality, the difference being that the ranking variable in this case is not health but rather socio-economic status. This aversion to socio-economic inequality would need taking into account into the social welfare function." (van Doorslaer et al., 1997, p. 96)

$$CI = \frac{2}{\mu}\sum_{t=1}^{T} f_t \mu_t R_t - 1$$
$$\mu = \sum_{t=1}^{T} f_t \mu_t$$
$$R_t = \sum_{\alpha=1}^{t-1} f_\alpha + \frac{1}{2} f_t$$

(1)

where μ is mean ill-health, f_t is the proportion of individuals in socio-economic group t, μ_t is ill-health in socio-economic group t and R_t is the relative rank of socio-economic group t. CI takes values between –1 and +1. It takes –1 (+1) if total ill-health is solely concentrated among the least (most) advantaged socio-economic group. Intermediate values are either favouring the more or less advantaged socio-economic groups *overall.* The italics are important since it is theoretically possible that the corresponding concentration curve crosses the diagonal.

Kakwani et al. (1997) constructed a distribution-free asymptotic estimator for the variance of CI. It can be used to perform hypothesis testing.

$$\mathrm{var}(CI) = \frac{1}{T}\left[\sum_{t=1}^{T} f_t a_t^2 - (1 + CI)^2\right]$$
$$a_t = \frac{\mu_t}{\mu}(2R_t - 1 - CI) + 2 - q_{t-1} - q_t$$
$$q_t = \frac{1}{\mu}\sum_{\alpha=1}^{t} f_\alpha \mu_\alpha$$

(2)

In the above description, I have assumed that we have a continuous ill-health measure. However, as in most data sets, I have a categorical ill-health measure at my disposal (e.g. how is your health status in general: very good, good, moderate, bad or very bad?). Therefore, one has to transform the categorical measure into a continuous one. Wagstaff & van Doorslaer (1994) propose a procedure based on the standard lognormal distribution. The standard lognormal distribution is chosen for since it imposes skewness in the distribution of ill-health, i.e. only a minority of the population claims to suffer from severe illness whereas the majority has no illness whatsoever. The continuous ill-health measure is obtained by dividing the area under the standard lognormal density function. It is divided up in such a way that the areas correspond to the sample proportions of each respond category. One then obtains a continuous measure that gives higher values to more severe ill-health statuses. The whole procedure, however, relies crucially upon the choice for the standard lognormal distribution. Since there is no other justification than the skewed nature of this distribution function, some authors have compared the use of the above constructed ill-health measure with other ill-health measures that are continuous by definition (e.g. time trade-off method, etc.). Gerdtham et al. (1999) and Humphries & van Doorslaer (2000) found no statistically significant difference between the various methods. It results that there is additional evidence that the standard lognormal distribution is a reasonable choice indeed.

Finally, some words on standardisation and on ill-health versus health. As Sen (1992) argued, there might be a difference between measuring inequality of shortfalls and inequality of attainments, i.e. between inequality of ill-health versus health. "Should a person's position be judged, positively, in terms of the level of achievement, or negatively, in terms of the shortfall vis-à-vis what she could have maximally achieved?" (cit., Sen, 1992) The answer is a difficult one, but it should be clear that both strategies could give rise to different conclusions if the maximally achievable position varies over the population. It might well be that all persons are in equal health, but only few would be willing to state that there is no inequality in this situation. A person of twenty years experiencing the same health as a person of eighty years seems unequal since the former may be expected to have a greater shortfall than the latter.

Because of the above I use inequality of ill-health instead of inequality of health throughout this paper. The distinction is however unclear since I use a subjective ill-health variable; it depends on whether individuals answer in terms of attainments or shortfalls on the "how is your health in general?" question. Moreover, defining what is a person's maximally achievable health status seems even a greater challenge. In the context of the ECuity project, one has addressed the problem by standardising the ill-health variable. In this paper, I use the direct standardisation method. "This involves applying the age-sex specific average illness rate of each socio-economic group to the age and gender structure of the population. The standardised ill-health rate for socio-economic group t is equal to:" (cit., Wagstaff et al., 1997)

$$\mu_t^+ = \sum_d \frac{n_d \mu_{dt}}{n} \qquad (3)$$

where n_d is the number of persons in d^{th} demographic group and μ_{dt} is the ill-health rate amongst persons in the d^{th} demographic group in socio-economic group t. If one would not make a demographic adjustment, one would assume that all inequality is avoidable. Remembering the example of the twenty year versus eighty year old person, it seems reasonable to assume that standardisation eliminates unavoidable inequality attributable to demographic differences over the socio-economic groups.

The procedure outlined in this section can be briefly summarised as follows. First, one has to transform the categorical ill-health measure into a continuous one. Next, one applies direct standardisation to the continuous ill-health measure in order to be able to measure avoidable inequality. Finally, a concentration curve or index is calculated based on the standardised ill-health measure.

2.2. Estimating permanent income

King & Dicks-Mireaux (1982) examined the life-cycle hypothesis for wealth holdings. This hypothesis assumes a non-linear relationship between the ratio of wealth to permanent income and age. They define permanent income as normal age-adjusted earnings. In the remainder of this section, I summarise their estimating strategy for permanent incomes.

Permanent income, Y_i, is defined as follows:

$$\begin{aligned} &\ln(Y_i) = X_i\beta + S_i - c(A_i) \\ &E(S_i) = 0 \text{ and } \operatorname{var}(S_i) = \sigma_s^2 \end{aligned} \tag{4}$$

where X_i is a vector of observable individual characteristics, S_i consists of a vector of unobservable individual characteristics and $c(A_i)$ denotes the cohort effect. The latter is included since "for given X_i younger generations are better off than their elders because of technical progress and capital accumulation" (cit., King and Dicks-Mireaux, 1982). We, however, do observe current earnings instead of permanent income. Current earnings, E_{it}, are assumed to deviate from permanent income because of an age-earnings profile and transitory components of earnings (which are assumed independent of S_i).

$$\begin{gathered}\ln(E_{it}) = X_i\beta + S_i - c(A_{it}) + f(A_{it} - \ddot{A}) + u_{it} \\ E(u_{it}) = 0 \text{ and } \operatorname{var}(u_{it}) = \sigma_u^2\end{gathered} \tag{5}$$

where f denotes the age-earnings profile, Ä is the age on which permanent income is evaluated and u_{it} denotes the transitory component of current earnings.

In fact, I will be using (5) to obtain estimates of permanent income (evaluated at age Ä). Since I use a cross-section, one cannot estimate S_i, nor can one separately identify the age-earnings profile and the cohort effect. King & Dicks-Mireaux propose the following strategies. To obtain an estimate of the age-earnings profile, f, they "assume that one-half of the growth of real earnings was accounted for by…factors used as explanatory variables in the earnings equation, and that the other half was accounted for by technological progress and capital accumulation. The latter is the cohort effect." (cit., King and Dicks-Mireaux, 1982) Estimates of S_i are obtained by applying ols to (5) and noting that:

$$E\left(S_i \middle| S_i + u_{it}\right) = \hat{S}_i = \delta\left(S_i + u_{it}\right) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_u^2}\left(S_i + u_{it}\right)$$

(6)

The use of a cross-section implies that we restrict the value of δ to a particular value. Since particular values are chosen for the importance of the cohort effect vis-à-vis the age-earnings profile and for δ, I perform various sensitivity analyses. Combining (4) – (6) we get an estimate for permanent income:

$$\ln\left(Y_i^e\right) = X_i\hat{\beta} + \hat{S}_i - c\left(A_i\right)$$

(7)

The estimate for permanent income is used as stratification variable to compute concentration indices.

3. Data

I use the PSBH data set, a sample of the Belgian population over 15 years. In 1994, 6461 individuals were questioned on different socio-economic variables, including health status.

In the introduction, I mentioned that previous but unpublished research pointed out that there does not seem to exist any income-related inequality in ill-health among the 65+. In table 1, the various variables for this specific analysis are shown.

[Table 1]

After deleting missing values, 5926 observations were left. Income is household income (defined as monthly net disposable income) corrected for family structure. Sah1-5 is a self-reported, categorical ill-health measure. Age and sex are used to standardise the ill-health measure. In line with the ECuity project, I defined eight demographic groups, i.e. four age groups applied to both sexes (16-34, 35-44, 45-64 & 65+).

Estimating permanent income resulted in an additional number of missing values - 5548 observations were left. Therefore, I represent in table 2 all the variables used to estimate permanent income and to compute the corresponding concentration indices. The variables from table 1 are depicted for the reduced sample as well. They are used to compute a 'reference' concentration index, against which one can evaluate the results for the concentration indices based on permanent income.[4] (income has only 5110 observations, since I am only able to provide 5110 estimates of permanent income although I have 5548 observations, see section 5 for more)

[Table 2]

Labinc represents monthly net individual labour income and will be used as the dependent variable in (5). Households headed by widows are deleted since no information is available on the income of the deceased husband. Educ1-5 are four dummies, reflecting educational status. Educ2, standing for lower secondary school, is excluded from the regression analysis. Next, two regional dummies are constructed

[4] Calculation of the 'reference' index based on individual labour income seems obvious at first sight. However, the high number of zero incomes makes the analysis impracticable, i.e. concentration indices are based on rankings of the individuals in terms of their income. Since the number of zero incomes amounts to more than 2000 observations, none of the 2000 can be ranked.

to reflect the federal structure of Belgium (Flanders is suppressed). Income from financial assets is reflected in finas. Adult, child and out designate family structure. Out is particularly important since it takes one if the respective household has children that are independent and living outdoors. These children are not reflected in the adult variable, but might be important for obtaining a consistent estimate of permanent income of the older population. Disc1-5 is an additional self-reported, categorical ill-health measure. It focuses on chronic illness and determines the degree of discomfort suffered from it. Disc0, taking one if one reports no chronic illness, is excluded. Finally, there is a set of dummies reflecting occupational structure (persons with no work and unskilled labourers are suppressed).

Finally, some words on omitted, but possibly relevant variables for the regression analysis of equation (5). First, it was impossible to find a variable reflecting the demand side of the labour market. Several possibilities (unemployment rates) were looked for, but none of the definitions remained unchanged over the relevant period. Second, equation (5) is estimated on the level of the individual. Therefore, the labour market participation of the other household members might influence the labour market participation decision (and the number of hours worked) of the individual. However, for the elder subsample, the data set does not give any indication on the number of persons working in the household or on the number of hours worked by the other household members. Neither does it indicate the labour income of other household members. This is to be expected since the data set examines current position of the individuals. As such, retired individuals do not report past labour market position (past occupation is reported). Since I am primarily interested in the elder subsample, a specific regression technique is used to address the problem. In the next section I describe this regression technique and define the model in terms of observables.

4. Additional econometric theory and the model in terms of observables

I intend to estimate equation (5). In terms of observables we get:

$$\ln(\mathrm{labinc}_i) = \beta_0 + \beta_1 \mathrm{educ}_i + \beta_2 \mathrm{region}_i + \beta_3 f(\mathrm{age}_i - 45) + \beta_4 \mathrm{finas}_i + \beta_5 \mathrm{occup}_i - c(\mathrm{age}_i) + S_i + u_i \tag{8}$$

where educ, region and occup stand for respectively educational, federal and occupational sets of dummies.

The age earnings profile, f(age_i-45), reflects the possible non-linear relationship and indicates that permanent income is evaluated at the age of 45.

The cohort effect, c(age_i), is included to allow for the influence of capital accumulation and technological progress on labour income. Since I use a cross section, I have to impose a value (θ) for the relative importance of the cohort effect vis-à-vis the other explanatory variables. In the regression analyses, I use 0.3, 0.5 and 0.7. The cohort effect was implemented as follows: (1) data on growth rates of nominal earnings and consumer price indices, obtained from Nationaal instituut voor de statistiek (1931-1995) en Institut national de statistique (1915-1930), gave average annual growth rates of real earnings, -0.26% before 1924, -0.57% in 1925-1934, -3.86% in 1935-1944, 7.69% in 1945-1954, 3.93% in 1955-1964, 5.8% in 1965-1974, -0.03% in 1975-1984 and 1.25% in 1985-1994. (2) c(age_i) is defined as a piecewise linear function that takes zero at the age of 45:

$$
\begin{array}{ll}
age_i \leq 25 & -\theta[0.580-0.003+0.0125(25-age_i)] \\
25 < age_i \leq 35 & -\theta[0.580-0.0003(35-age_i)] \\
35 < age_i \leq 45 & -\theta[0.0580(45-age_i)] \\
45 < age_i \leq 55 & +\theta[0.0393(age_i-45)] \\
55 < age_i \leq 65 & +\theta[0.393+0.0769(age_i-55)] \\
65 < age_i \leq 75 & +\theta[0.393+0.769-0.0386(age_i-65)] \\
75 < age_i \leq 85 & +\theta[0.393+0.769-0.386-0.0057(age_i-75)] \\
85 < age_i & +\theta[0.393+0.769-0.386-0.057-0.0026(age_i-85)]
\end{array}
$$

(9)

In section 2.2, I explained one has to restrict δ to a particular value (see equation (6)) to obtain estimates for S_i. It should be clear that the value is influenced by the number of explanatory variables included in (5) and (8), i.e. the more explanatory variables, the lower the value. King & Dicks-Mireaux (1982) impose 0.5. Their set of explanatory variables is, however, slightly larger than mine. Therefore, I use 0.3, 0.5 and 0.7.

A second but more important caveat concerns the empirical content of S_i, defined as unobservable individual characteristics. Given the data limitations for the elder subsample (the data set only informs on past occupation), it may include more than unobservable individual characteristics. Since past hours of work are not given for the retired, hours of work are not included in the vector of explanatory variables, and are thus reflected in S_i. "The individual effect, S_i, may include variations in individual tastes for leisure, but it does not allow for systematic changes in annual hours worked resulting from spells of unemployment during part of the year, temporary lay-offs, or part-time work by wives or in retirement." (cit., King and Dicks-Mireaux, 1982) The solution is to estimate equation (8) on a truncated sample and to use the estimated parameters to predict permanent income for the entire sample. Such a procedure would however result in sample selection bias. The issue is addressed by means of the Heckit estimator (e.g. Greene, 1981, 2000, Heckman, 1979 and Melino, 1982).

The Heckit estimator consists of a three-step procedure.

In the first step, one estimates a probit model on the entire sample to determine the probability to be in the truncated sample. Since hours of work induce the problem of sample selection bias, I define a binary variable (trunc) that takes one (zero) if an individual works more than (less, including not working and retirement) 34 hours a week.[5] In fact, I estimate the following probit model on the entire sample:

$$\begin{aligned} \mathrm{trunc}_i &= \alpha_0 + \alpha_1 \mathrm{educ}_i + \alpha_2 \mathrm{region}_i + \alpha_3 f(\mathrm{age}_i - 45) + \alpha_4 \mathrm{finas}_i \\ &+ \alpha_5 \mathrm{hh}_i + \alpha_6 \mathrm{sah}_i + \alpha_7 \mathrm{disc}_i \end{aligned} \qquad (10)$$

where hh_i reflects the household structure and sah_i, $disc_i$ are sets of morbidity dummies.

[5] Employees claiming to work more than 34 hours a week but earning less than 30000 BEF, are deleted.

In the second step one runs ols on equation (8) for the truncated sample (trunc equals one), but the inverse mill's ratio is included as an additional explanatory variable. Finally, the parameter estimates obtained from the two steps are used as starting values for maximum likelihood estimation.

Another consequence of the omission of hours of work concerns the estimation of permanent income. First, we impose that $\hat{S}_i$ equals zero for those individuals not in the truncated sample. Second, equation (7) is changed to adjust explicitly for non-participation in the labour force.

$$Y_i^{e'} = Y_i^e \text{prob}(\text{trunc} = 1) + \overline{E}_i \text{prob}(\text{trunc} = 0) \qquad (11)$$

where the probabilities are obtained from equation (10), and $\bar{E}_i$ denotes the average of individual labour income of those working less than 34 hours.

5. Results

In section 5.1, I briefly describe the results for inequality in ill-health using income as stratification variable. The estimates for permanent income and the resulting concentration indices are discussed in section 5.2.

5.1. Income-related inequality in ill-health

As mentioned in the introduction, there isn't any income-related inequality in ill-health among the 65+ in Belgium. In table 3, I present the evidence.

[Table 3]

I used the variables from table 1 for the analysis. The concentration indices were calculated using grouped data, i.e. ten socio-economic groups (defined as deciles). Since calculation based on grouped data nearly differed from calculation based on individual level data [CI(full sample)=-0.89, CI(65-)=-0.114, CI(65+)=-0.032] and since comparative research within the ECuity project uses grouped data (e.g. van Doorslaer et al., 1997), I use grouped data throughout the paper.

The results can be summarised as follows: First, there is significant socio-economic inequality in ill-health, favouring the rich, among the whole population and among the 65-. Second, there is no statistical significant socio-economic inequality in ill-health among the 65+. Third, socio-economic inequality in ill-health among the 65- is statistically significant higher than among the 65+.

5.2. Permanent income-related inequality in ill-health

In this section, I first describe the estimates obtained by the Heckit estimator. Next, I describe the estimates of permanent income, evaluated at the age of 45 and finally, the concentration indices, based on permanent income are dealt with.

In table 4, the estimation results are depicted.[6] In this estimation strategy, I imposed that one-half of the growth of real earnings is attributable to the cohort effect. In the appendix, the estimation results for cohort effects of 30% and 70% are given. I estimated separately for males and females, since joint estimation resulted in heteroskedasticity.

[Table 4]

The selection equation determines the probability to be in the truncated sample. For both males and females, higher education results in a higher probability to be working more than 34 hours a week. The relationship between age and working more than 34 hours a week is highly non-linear (a fifth order polynomial fitted the data best), but indicates that the younger and older are less likely to be in the truncated sample. Income from financial assets has a positive influence for males (not for females). It could mean that income from financial assets is a proxy for unobservable characteristics. However, it might well be the case that a higher number of hours worked increases one's opportunity to build up financial assets (e.g., Miles, 1997). Family structure is not important for males. Females are less likely to work more than 34 hours if the number of household members is larger. Finally, I included two ill-health variables in the selection equation. The coefficients indicate that males suffering from more severe illness are less likely to be in the truncated sample. The inclusion of ill-health variables might appear strange, given that I ultimately want to examine the relationship between income and ill-health status. However, no ill-health

[6] Trimming for 1 % outliers did not change the results qualitatively.

variables were included in the earnings equation and the ill-health variables thus only pick up effects of sample selection bias.

Next, I turn to the earnings equation. Since most explanatory variables are represented by dummies, one can interpret the coefficients as percentage changes (against the suppressed dummy) in earnings. First, higher education results in higher earnings. Second, the age-earnings profile is linear. Finally, the occupational dummies reflect that a male professional, farmer, company manager and executive has higher earnings than an unskilled labourer has. This pattern is, however, less clear for females.

The estimated coefficients are used to predict permanent income. Equation (6) shows that $\hat{S}_i$ equals $\delta(S_i + u_i)$ for the observations in the truncated sample. I impose that δ equals 0.5. Since $\hat{S}_i$ cannot be estimated for individuals not in the truncated sample, those individuals get $\hat{S}_i$ equals zero. Next, equation (7) and (11) are applied on both the male and female sample to obtain *individual* permanent incomes. If one would use these permanent incomes to calculate concentration indices, one faces two problems. First, the 'reference' concentration index is calculated based on equivalent income (see income in table 2). Second and more importantly, one assumes that estimated *individual* permanent income reflects the socio-economic status of the individual. This implies for example that a male with a high permanent income is a member of an advantaged socio-economic group, whereas his inactive wife belongs to a disadvantaged socio-economic group. I address both problems by adding up the *individual* permanent incomes within each household to obtain *household* permanent income. This procedure resulted in an additional number of missing values. An individual may be included in the regression analysis, but some of the other household members may be excluded. Deleting those households reduced the sample from 5548 to 5110 observations. In the following step, *household* permanent income is corrected for family structure, using (adults+0.5children)$^{0.5}$ as equivalence scale. Since permanent income is evaluated at the age of 45 years, the number of children denotes the number the individual has when he was/is/will be 45 years. Finally, I calculated concentration indices based on equivalent permanent income.

In table 5, estimated equivalent permanent income, the 'reference' concentration indices and concentration indices based on permanent income are given. In the appendix, the results of various sensitivity analyses are shown.

[Table 5]

Estimated equivalent permanent income was obtained by applying the above-described procedure (cohort effect equals 0.5, $\hat{S}_i = 0.5(S_i + u_i)$ for persons in the truncated sample and non-participation in the labour force is adjusted for).
The 'reference' concentration indices are based on 5110 observations of *equivalent income* (see table 2). Comparing them with the concentration indices of table 3, we observe that all indices slightly increase. Apparently the reduction in the number of observations, required to make meaningful comparisons with the concentration indices based on permanent income, results in a slightly higher socio-economic inequality in ill-health. Turning towards statistical significance, the conclusions nearly change. There is still significant socio-economic inequality in ill-health, favouring the rich, among the entire population and the 65-. Moreover, the difference in socio-economic inequality in ill-health between the 65- and 65+ still holds. Only

with respect to the socio-economic inequality in ill-health among the 65+, the conclusions change, i.e. we have pro-rich inequality. However, a simple t-test shows that the CI(65+) does not significantly differ from –0.005. It is to be hoped that the limited deviation of the concentrations indices based on the reduced sample (table 5) from the concentration indices based on the original sample (table 3), indicates that the reduction of observations was random.

The concentration indices based on *equivalent permanent income* differ from those based on equivalent income. Nevertheless, the basic conclusions remain unchanged. The estimation of permanent income, however, invokes several assumptions. Therefore, the results of alternative estimation strategies are depicted in the appendix (table A.3). It is seen that the results are most sensitive to the imposed cohort effect. The higher the imposed cohort effect, the more significant is the difference in socio-economic inequality in ill-health between the 65- and 65+. In other words, the more important capital accumulation and technological progress are thought of for the explanation of the growth of real earnings, the higher (and the more significant) is the difference between the concentration indices of the 65- and 65+.[7]

A last remark concerns the appropriateness of using self-assessed health. Its use seems more or less justified in the context of equivalent income ('reference' concentration index) since ill-health as well as income are measures of current individual status. Permanent income, however, is not necessarily a characteristic of one's current status. Therefore, one could argue to use a ' permanent' or 'life-cycle' ill-

[7] In seeking an explanation, I thought of the following: a higher cohort effect reduces, ceteris paribus, the importance of the age-earnings profile (see equation (5)). The coefficients on the individual characteristics, however, should remain unchanged (and they do not differ significantly, see table 4, A.1 & A.2). As such the difference in predicted permanent income (see equation (7)) should stem from the unobservable individual characteristics or from the cohort effect. I checked whether the unobservables differ significantly across the imposed cohort effects and they did not. As such, we have, and this is a tentative explanation indeed, an indication that the imposition of higher cohort effects results in increased divergence between the predicted permanent incomes of the elderly and the

health measure. Mortality is frequently used within this context. Work on the relationship between mortality and self-assessed health, however, has indicated that self-assessed health is an important predictor of mortality (e.g., Idler and Angel, 1990 and Kaplan and Camacho, 1983).

Another way to check the robustness of the results, consists of calculating concentration indices based on other measures of socio-economic status. Since I am primarily interested in the 65+ and since I have to take into account the limitations of the data set, I choose rental value of houses. The (rental) value of houses might be a good indicator of the socio-economic status of the elderly. Accumulation of income and (wealth) may be reflected in the value of houses indeed, yet it is far from a perfect indicator. Does a household living in a large, high-valued house (e.g. a country-house), thereby possibly incurring high maintenance costs, attains a higher socio-economic status than a household living in a small, slightly lower valued house? Other examples can be thought of.
Additional problems arise because some households are proprietors and others are tenants. The tenants answer the following question: "How much rent do you pay for your house?" The proprietors face another question: "How much rent would you have to pay if you didn't own the house?" First, misreporting and problems encountered if measuring stated preference, might be more crucial for the proprietors. They are probably less accustomed with the current values in the rental market than tenants are. Moreover, strategic motives might result in conscious over- or understatement of the rental value of the house. Second, tenants actually have to pay the rental value of the house, whereas proprietors do not have to. For younger households, the problem may

---

younger. This could lead to a more significant difference in socio-economic inequality in ill-health between both subsamples.

not be that compelling, since proprietors probably are paying off a 'housing' loan. For the elderly this argument is less convincing and thus points to possible incomparability of tenants and proprietors. I address the problem by means of the following procedure: First, I calculate concentration indices for the entire population, i.e. tenants and proprietors. Next, the focus is on the subsamples. In table 6, the concentration indices based on rental values are depicted. Note that once again I use the same equivalence scale to correct the rental value of houses for family structure, i.e. $(\text{adults}+0.5\text{children})^{0.5}$.

[Table 6]

It is immediately seen that the incorporation of rental values points to a somewhat different pattern than the analysis based on income and permanent income. There is still significant socio-economic inequality in ill-health among the 65- and among the full sample, but the difference between 65- and 65+ is only statistically significant if one restricts the attention to tenants. Two conclusions could be drawn, but it is not clear which of both is to be preferred. First, one could conclude that the distinction between the 65- and 65+ is not existing. However, this implicitly implies that the rental value of houses is a better indicator of one's socio-economic status than (permanent) income. Second, one may conclude that evidence on differences in socio-economic inequality in ill-health between the 65- and 65+ is at best mixed. Further research should focus on other measures of socio-economic status and the analysis could be replicated for other countries to test the stated hypothesis, i.e. socio-economic inequality in ill-health among the 65+ is limited or even not existing.[8]

[8] Another possibility for future research consists of using decomposable measures. They do not suffer from the 're-ranking' problem. Moreover, they can be used to determine the relative contribution of the inequality in a subgroup to overall inequality. However, the current state of the art, as far as I know,

6. Conclusions

In this paper, I focused on socio-economic inequality in ill-health, i.e. are the poorer members in better (worse) health than the richer members of society are? This research question is closely connected with the research undertaken within the ECuity project. Since the ECuity project undertakes comparative research of health care systems across European countries, a specific methodology to quantify (and thus to compare) socio-economic inequality in ill-health was developed. First, the population is ranked in terms of socio-economic status. Second, concentration curves and indices are calculated to quantify the inequality.

The specificity of this paper concerns the use of permanent income as stratification variable instead of income. I opted to use permanent income since there doesn't seem to exist any *income*-related inequality in ill-health among the 65+. The variation in income among the 65+ is rather limited (because of welfare benefits) and it is thus not a priori clear whether the absence of inequality stems form the limited variation in income (assuming that the variation is underestimated, e.g. income from financial assets, etc.) or truly reflects reality. The estimation strategy of permanent income was based on the methodology proposed by King & Dicks-Mireaux (1982).

In the empirical part of the paper, I used Belgian data for 1994.

First, I replicated the analysis already undertaken within the ECuity project and complemented it with analyses of subsamples of the Belgian population, i.e. the 65-

---

makes them useless since they cannot be applied to two attributes (being socio-economic status and ill-health) simultaneously.

and 65+. I found that there is significant *income*-related inequality in ill-health in Belgium, favouring the rich. The analysis of the subsamples pointed out that 1) income-related inequality in ill-health among the 65- is favouring the rich, 2) it is absent for the 65+ and 3) inequality among the 65- is significant higher than among the 65+.

Second, I calculated concentration indices based on *permanent income* for the same (sub)sample(s). The basic conclusions remained unchanged (indicating that the limited variation in the income of the 65+ is not driving the result). However, one should be cautious in interpreting the results too literally, because the estimation of permanent income invoked several assumptions.

To check the robustness of the results, I did the analysis over again using *rental value of houses* as a measure of socio-economic status. The difference between the 65- and 65+ turned out to be insignificant.

In my opinion, there is evidence on the absence of socio-economic inequality in ill-health among the 65+ indeed, yet the evidence is fragile and further exploration is needed.

Appendix

Various sensitivity analyses of the estimation of permanent income are given in this appendix. Table A.1 and A.2 depict the estimation results of the Heckit estimator for cohort effects of 30% and 70%.

[Table A.1]

[Table A.2]

Table A.3 gives an overview of concentration indices based on different estimation strategies for permanent income.

[Table A.3]

References

De Graeve, D. and I. Duchesne, 1997, Socio-economic health inequalities in Belgium in an international perspective, Arch Public Health 55, 119-138.

Gerdtham, U.-G., M. Johannesson, L. Lundberg and D. Isacson, 1999, A note on validating Wagstaff and van Doorslaer's health measure in the analysis of inequalities in health, Journal of health economics 18, 117-124.

Greene, W.H., 1981, Sample selection bias as a specification error: comment, Econometrica 49, 795-798.

Greene, W.H., 2000, Econometric analysis: fourth edition (Prentice Hall, New Jersey).

Hall, R.E. and F.S. Mishkin, 1982, The sensitivity of consumption to transitory income: estimates from panel data on households, Econometrica 50, 461-481.

Heckman, J.J., 1979, Sample selection bias as a specification error, Econometrica 47, 153-161.

Humphries, K.H. and E. van Doorslaer, 2000, Income-related health inequality in Canada, Social Science and medicine, forthcoming.

Idler, E.L. and R.J. Angel, 1990, Self-rated health and mortality in the NHANES-I epidemiologic follow-up study, American journal of public health 80, 446-452.

Institut National de Statistique, 1915-1930, Annuaire statistique de la Belgique et du Congo belge (Brussels).

Jappelli, T., 1999, The age-wealth profile and the life-cycle hypothesis: a cohort analysis with a time series of cross-sections of Italian households, Review of income and wealth 45, 57-75.

Kakwani, N., A. Wagstaff and E. van Doorslaer, 1997, Socio-economic inequalities in health: measurement, computation, and statistical inference, Journal of econometrics 77, 87-103.

Kaplan, G.A. and T. Camacho, 1990, Perceived health and mortality: a nine-year follow-up of the human population, American journal of epidemiology 117, 292-304.

King, M.A. and L-D. L. Dicks-Mireaux, 1982, Assets holdings and the life-cycle, The economic journal 92, 247-267.

Melino, A., 1982, Testing for sample selection bias, Review of economic studies 49, 151-153.

Miles, D., 1997, A household level study of the determinants of incomes and consumption, The economic journal 107, 1-25.

Nationaal Instituut voor de Statistiek, 1931-1959, Statistisch jaarboek voor België en Belgisch-Congo (Brussels).

Nationaal Instituut voor de Statistiek, 1960-1995, Statistisch jaarboek van België (Brussels).

Sen, A., 1992, Inequality reexamined (Clarendon Press, NY-Oxford).

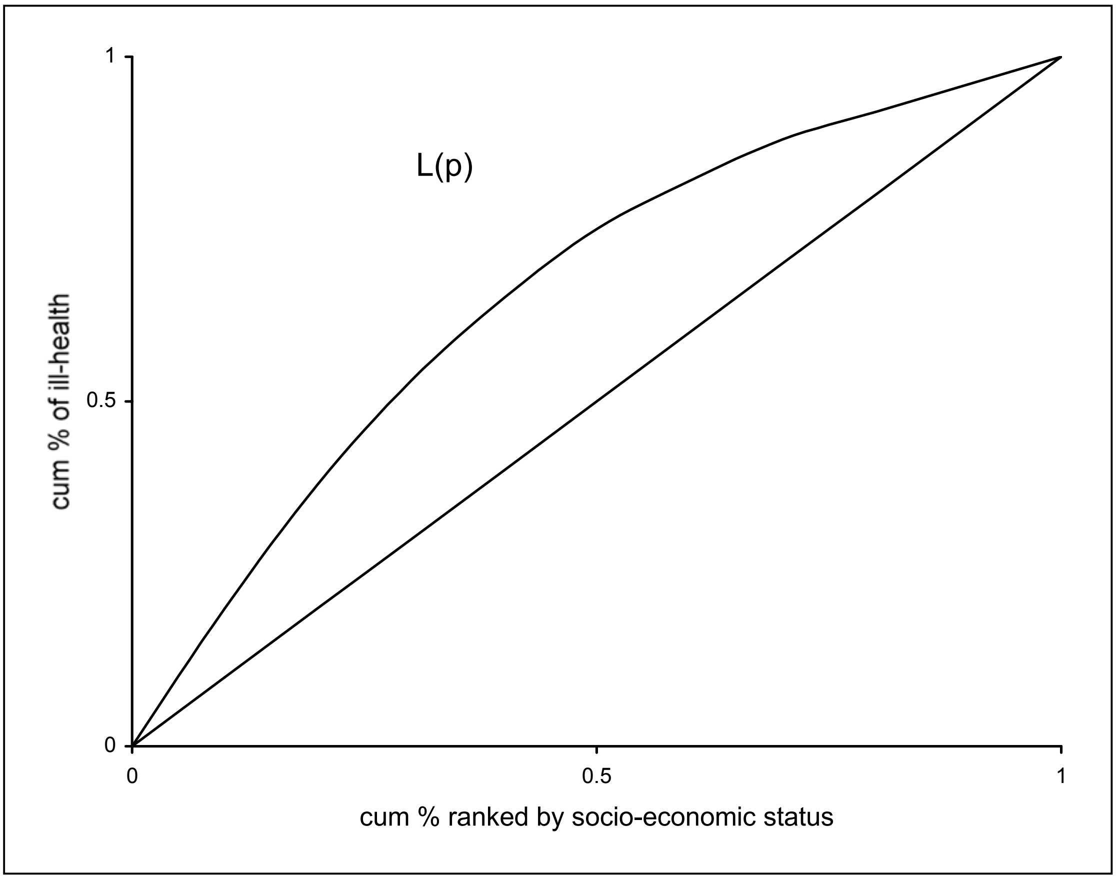
van Doorslaer, E. et al., 1997, Income-related inequalities in health: some international comparisons, Journal of health economics 16, 93-112.

Wagstaff, A., P. Paci and E. van Doorslaer, 1991, On the measurement of inequalities in health, Social science and medicine 33, 545-557.

Wagstaff, A. and E. van Doorslaer, 1994, Measuring inequalities in health in the presence of multiple-category morbidity indicators, Health economics 3, 281-291.

Willis, R.J., 1986, Wage determinants: a survey and reinterpretation of human capital earnings functions, in: D.C. Ashenfelter and R. Layard, eds., Handbook of labour economics, vol. 1 (North Holland, Amsterdam) 525-602.

Figure 1: illness concentration curve: L(p)



Source: De Graeve and Duchesne, 1997

Table 1: descriptive variables: *income*-related inequality

	obs	mean	stdev	description
income	5926	48312	24521	equivalent income: household income/(adults+0.5*children)$^{0.5}$
sah1	5926	0.24	0.43	self assessed health: how is your health in general? very good
sah2	5926	0.49	0.50	self assessed health: how is your health in general? good
sah3	5926	0.22	0.42	self assessed health: how is your health in general? moderate
sah4	5926	0.04	0.20	self assessed health: how is your health in general? bad
sah5	5926	0.01	0.09	self assessed health: how is your health in general? very bad
sex	5926	0.48	0.50	1=male; 0=female
age	5926	46.50	18.49	

Table 2: descriptive variables: *permanent income*-related inequality

	obs	mean	stdev	description
income	5110	48127	24200	equivalent income: household income/ (adults+0.5*children)$^{0.5}$
sah1	5548	0.25	0.43	self assessed health: how is your health in general? very good
sah2	5548	0.49	0.50	self assessed health: how is your health in general? good
sah3	5548	0.21	0.041	self assessed health: how is your health in general? moderate
sah4	5548	0.04	0.19	self assessed health: how is your health in general? bad
sah5	5548	0.01	0.08	self assessed health: how is your health in general? very bad
sex	5548	0.48	0.50	1=male; 0=female
age	5548	44.01	17.08	
labinc	5548	35519	136072	net individual labour income
educ1	5548	0.16	0.37	no or primary education: 1=yes; 0=no
educ3	5548	0.29	0.45	higher secondary school: 1=yes; 0=no
educ4	5548	0.31	0.46	higher education: 1=yes; 0=no
educ5	5548	0.02	0.14	graduate-level education: 1=yes; 0=no
wal	5548	0.44	0.50	inhabitant: 1=Wallonia; 0=otherwise
brus	5548	0.12	0.32	inhabitant: 1=Brussels; 0=otherwise
finas	5548	0.20	0.40	income from financial assets: 1=yes; 0=otherwise
adult	5548	3.09	1.33	number of adults
child	5548	0.71	1.02	number of children (under 16 years)
out	5548	0.28	0.45	children - outdoors & independent: 1=yes; 0=otherwise
disc1	5548	0.02	0.13	chronic illness: 1=maximum discomfort; 0=otherwise
disc2	5548	0.04	0.19	chronic illness: 1=severe discomfort; 0=otherwise
disc3	5548	0.07	0.25	chronic illness: 1=moderate discomfort
disc4	5548	0.02	0.15	chronic illness: 1=limited discomfort; 0=otherwise
disc5	5548	0.02	0.13	chronic illness: 1=no discomfort; 0=otherwise
self	5548	0.04	0.20	self-employed: 1=yes; 0=otherwise
farm	5548	0.01	0.09	farmer: 1=yes; 0=otherwise
prof	5548	0.02	0.13	professionals (lawyers, GP, etc.): 1=yes; 0=otherwise
comp	5548	0.004	0.07	company manager: 1=yes; 0=otherwise
whol	5548	0.001	0.03	wholesale dealer: 1=yes; 0=otherwise
selfr	5548	0.01	0.11	self-employed - residual category: 1=yes; 0=otherwise
skil	5548	0.13	0.33	skilled worker: 1=yes; 0=otherwise
emp	5548	0.26	0.44	employee: 1=yes; 0=otherwise
exec	5548	0.11	0.31	executive: 1=yes; 0=otherwise
empr	5548	0.03	0.16	employee – residual category: 1=yes; 0=otherwise

Table 3: income-related inequality in ill-health in Belgium in 1994

	full sample	65-	65+
obs	5926	4678	1248
CI	-0.090*	-0.112*	-0.031
se (CI)	0.0122	0.0225	0.0185
hypothesis testing (t-values)			
65-			-2.78

*: significant at 5%

Table 4: maximum likelihood estimation (cohort effect: θ = 0.5)

probability to be in the truncated sample: dependent variable is trunc				
	males		**females**	
	coefficient	standard error	coefficient	standard error
educ1	-0.322*	0.123	-0.184	0.134
educ3	0.066	0.095	0.321*	0.092
educ4	0.117	0.048	0.390*	0.094
educ5	0.492*	0.093	0.626*	0.228
wal	-0.239*	0.072	-0.139	0.071
brus	-0.370*	0.106	0.235	0.102
age-45	1.470*	0.283	1.228*	0.374
(age^2)-(45^2)	-0.045*	0.011	-0.042*	0.015
finas	0.433*	0.089	0.230	0.094
adult	0.061	0.035	-0.177*	0.039
child	-0.116	0.049	-0.190*	0.046
out	-0.084	0.098	-0.262	0.122
sah2	0.033	0.078	-0.052	0.076
sah3	-0.291*	0.107	-0.216	0.103
sah4	-1.230*	0.258	-0.452	0.280
sah5	-1.141*	0.228	-0.448	0.187
disc1	-0.479*	0.160	-0.291	0.118
disc2	-0.612*	0.232	-0.558	0.342
disc3	-0.334	0.143	-0.433	0.175
disc4	-0.127	0.216	-0.274	0.347
disc5	0.062	0.260	0.277	0.310
estimation of earnings equation: dependent variable is ln(labinc)				
educ1	-0.088	0.069	-0.139	0.320
educ3	0.069	0.044	0.101	0.074
educ4	0.193*	0.048	0.208*	0.079
educ5	0.273*	0.083	0.318*	0.081
wal	0.031	0.032	-0.049	0.054
brus	0.015	0.049	-0.004	0.072
age-45	0.027*	0.002	0.027*	0.004
finas	0.080	0.035	0.100	0.055
self	-0.036	0.083	0.037	0.185
farm	0.587*	0.093		
prof	0.419*	0.107	0.681*	0.198
comp	0.458*	0.094	0.892*	0.270
whol	-0.176	0.244	-2.347	4.138
selfr	0.920*	0.156	-0.719	4.179
skil	0.009	0.085	-0.047	0.282
emp	0.001	0.091	0.125	0.185
exec	0.232*	0.092	0.294	0.201
empr	-0.134	0.416	-0.088	0.757
obs	2709		2829	
log likelihood	-2056		-1531	

Constants are suppressed; in the probability to be in the truncated sample, f(age-45) was approximated by a fifth order polynomial; farm is excluded since no female claims to be farmer; *: significant at 1%

Table 5: descriptive variables of permanent income and *permanent income*-related versus *income*-related inequality in ill-health in Belgium in 1994

descriptive variables				
	obs	**mean**	**stdev**	**description**
perm	5110	37661	81107	equivalent permanent income

'reference' concentration indices			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.101*	-0.118*	-0.035*
se (CI)	0.0151	0.0208	0.0178
hypothesis testing (t-values)			
65-			-3.03
concentration indices based on equivalent permanent income			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.112*	-0.117*	-0.041
se (CI)	0.0169	0.0212	0.0229
hypothesis testing (t-values)			
65-			-2.44

*: significant at 5%. Permanent income (perm) was calculated assuming the following: 1) cohort effect: 0.5, 2) $\hat{S}_i$ equals $0.5(S_i + u_i)$ for the individuals in the truncated sample and equals zero for individuals not in the truncated sample, 3) equation (7) and (11) are used. The 'reference' concentration index is based on equivalent income.

Table 6: *housing value*-related inequality in ill-health in Belgium in 1994

concentration indices for tenants and proprietors			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.091*	-0.103*	-0.064*
se (CI)	0.0121	0.0180	0.0205
hypothesis testing (t-values)			
65-			-1.43
concentration indices for proprietors			
	full sample	**65-**	**65+**
obs	3830	3254	576
CI	-0.052*	-0.054*	-0.042
se (CI)	0.0090	0.0059	0.0204
hypothesis testing (t-values)			
65-			-0.57
concentration indices for tenants			
	full sample	**65-**	**65+**
obs	1280	1106	174
CI	-0.071*	-0.098*	0.013
se (CI)	0.0278	0.0203	0.0509
hypothesis testing (t-values)			
65-			-2.03

*: significant at 5%.

Table A.1: maximum likelihood estimation (cohort effect: θ = 0.3)

probability to be in the truncated sample: dependent variable is trunc				
	males		**females**	
	coefficient	standard error	coefficient	standard error
educ1	-0.322*	0.123	-0.184	0.139
educ3	0.066	0.095	0.321*	0.093
educ4	0.117	0.048	0.390*	0.096
educ5	0.492*	0.093	0.626*	0.227
wal	-0.239*	0.072	-0.139	0.072
brus	-0.370*	0.106	0.235	0.103
age-45	1.470*	0.282	1.228*	0.368
(age^2)-(45^2)	-0.045*	0.011	-0.042*	0.015
finas	0.433*	0.089	0.230	0.094
adult	0.061	0.035	-0.177*	0.039
child	-0.116	0.049	-0.191*	0.046
out	-0.084	0.097	-0.262	0.122
sah2	0.033	0.077	-0.052	0.076
sah3	-0.291*	0.105	-0.216	0.103
sah4	-1.230*	0.255	-0.452	0.280
sah5	-1.141*	0.226	-0.448	0.188
disc1	-0.479*	0.154	-0.291	0.120
disc2	-0.612*	0.225	-0.558	0.341
disc3	-0.334	0.139	-0.433	0.173
disc4	-0.127	0.214	-0.274	0.349
disc5	0.063	0.257	0.277	0.310
estimation of earnings equation: dependent variable is ln(labinc)				
educ1	-0.080	0.068	-0.133	0.297
educ3	0.066	0.044	0.091	0.074
educ4	0.202*	0.048	0.196*	0.072
educ5	0.272*	0.083	0.309*	0.081
wal	0.035	0.032	-0.043	0.054
brus	0.023	0.049	-0.006	0.072
age-45	0.020*	0.002	0.021*	0.004
finas	0.071	0.035	0.094	0.055
self	-0.033	0.080	0.040	0.183
farm	0.586*	0.088		
prof	0.424*	0.104	0.685*	0.196
comp	0.463*	0.091	0.883*	0.271
whol	-0.127	0.252	-2.362	4.141
selfr	0.932*	0.151	-0.729	4.188
skil	0.010	0.083	-0.048	0.281
emp	0.004	0.087	0.128	0.183
exec	0.234*	0.090	0.296	0.198
empr	-0.109	0.349	-0.086	0.784
obs	2709		2839	
log likelihood	-2078		-1533	

Constants are suppressed; in the probability to be in the truncated sample, f(age-45) was approximated by a fifth order polynomial; farm is excluded since no female claims to be farmer; *: significant at 1%

Table A.2: maximum likelihood estimation (cohort effect: $\theta = 0.7$)

	males		females	
	coefficient	standard error	coefficient	standard error
probability to be in the truncated sample: dependent variable is trunc				
educ1	-0.322*	0.122	-0.184	0.134
educ3	0.066	0.095	0.321*	0.092
educ4	0.117	0.048	0.390*	0.094
educ5	0.492*	0.093	0.626*	0.228
wal	-0.239*	0.072	-0.139	0.071
brus	-0.370*	0.106	0.235	0.102
age-45	1.470*	0.282	1.228*	0.376
(age^2)-(45^2)	-0.045*	0.010	-0.042*	0.015
finas	0.433*	0.089	0.230	0.094
adult	0.061	0.036	-0.177*	0.039
child	-0.116	0.050	-0.191*	0.046
out	-0.084	0.098	-0.262	0.122
sah2	0.033	0.078	-0.052	0.076
sah3	-0.291*	0.107	-0.216	0.103
sah4	-1.230*	0.258	-0.452	0.281
sah5	-1.141*	0.226	-0.449	0.189
disc1	-0.479*	0.154	-0.291	0.120
disc2	-0.612*	0.233	-0.558	0.339
disc3	-0.334	0.142	-0.433	0.175
disc4	-0.127	0.214	-0.274	0.344
disc5	0.063	0.261	0.277	0.309
estimation of earnings equation: dependent variable is ln(labinc)				
educ1	-0.095	0.069	-0.144	0.334
educ3	0.072	0.044	0.111	0.074
educ4	0.185*	0.049	0.220*	0.075
educ5	0.275*	0.084	0.328*	0.081
wal	0.026	0.032	-0.056	0.055
brus	0.006	0.049	-0.002	0.073
age-45	0.034*	0.002	0.033*	0.004
finas	0.089	0.036	0.107	0.054
self	-0.040	0.083	0.035	0.181
farm	0.589*	0.094		
prof	0.414*	0.108	0.676*	0.194
comp	0.453*	0.094	0.901*	0.265
whol	-0.225	0.236	-2.333	4.133
selfr	0.908*	0.159	-0.708	4.170
skil	0.007	0.085	-0.045	0.273
emp	0.001	0.091	0.123	0.181
exec	0.229*	0.090	0.292	0.197
empr	-0.159	0.445	-0.090	0.699
obs	2709		2839	
log likelihood	-2061		-1533	

Constants are suppressed; in the probability to be in the truncated sample, f(age-45) was approximated by a fifth order polynomial; farm is excluded since no female claims to be farmer; *: significant at 1%

Table A.3: concentration indices based on permanent income: sensitivity analyses

CI based on permanent income: *1) cohort effect: 0.3, 2) $\hat{S}_i$ equals 0.3(S_i + u_i) for the individuals in the truncated sample and equals zero for individuals not in the truncated sample, 3) equation (7) and (11) are used.*			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.111*	-0.126*	-0.070*
se (CI)	0.0200	0.0318	0.0279
hypothesis testing (t-values)			
65-			-1.32

CI based on permanent income: *1) cohort effect: 0.5, 2) $\hat{S}_i$ equals 0.3(S_i + u_i) for the individuals in the truncated sample and equals zero for individuals not in the truncated sample, 3) equation (7) and (11) are used.*			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.126*	-0.133*	-0.041
se (CI)	0.0258	0.0320	0.0228
hypothesis testing (t-values)			
65-			-2.34

CI based on permanent income: *1) cohort effect: 0.7, 2) $\hat{S}_i$ equals 0.3(S_i + u_i) for the individuals in the truncated sample and equals zero for individuals not in the truncated sample, 3) equation (7) and (11) are used.*			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.146*	-0.143*	-0.013
se (CI)	0.0241	0.0313	0.0197
hypothesis testing (t-values)			
65-			-3.52

CI based on permanent income: *1) cohort effect: 0.3, 2) $\hat{S}_i$ equals 0.5(S_i + u_i) for the individuals in the truncated sample and equals zero for individuals not in the truncated sample, 3) equation (7) and (11) are used.*			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.107*	-0.119*	-0.070*
se (CI)	0.0184	0.0262	0.0277
hypothesis testing (t-values)			
65-			-1.29

CI based on permanent income: *1) cohort effect: 0.7, 2) $\hat{S}_i$ equals 0.5(S_i + u_i) for the individuals in the truncated sample and equals zero for individuals not in the truncated sample, 3) equation (7) and (11) are used.*			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.120*	-0.128*	-0.014
se (CI)	0.0129	0.0216	0.0199
hypothesis testing (t-values)			
65-			-3.88

*: significant at 5%.

Table A.3: continued

CI based on permanent income: *1) cohort effect: 0.3, 2) $\hat{S}_i$ equals $0.7(S_i + u_i)$ for the individuals in the truncated sample and equals zero for individuals not in the truncated sample, 3) equation (7) and (11) are used.*			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.105*	-0.115*	-0.068*
se (CI)	0.0180	0.0260	0.0272
hypothesis testing (t-values)			
65-			-1.25

CI based on permanent income: *1) cohort effect: 0.5, 2) $\hat{S}_i$ equals $0.7(S_i + u_i)$ for the individuals in the truncated sample and equals zero for individuals not in the truncated sample, 3) equation (7) and (11) are used.*			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.106*	-0.111*	-0.041
se (CI)	0.0143	0.0177	0.0226
hypothesis testing (t-values)			
65-			-2.44

CI based on permanent income: *1) cohort effect: 0.7, 2) $\hat{S}_i$ equals $0.7(S_i + u_i)$ for the individuals in the truncated sample and equals zero for individuals not in the truncated sample, 3) equation (7) and (11) are used.*			
	full sample	**65-**	**65+**
obs	5110	4358	752
CI	-0.104*	-0.120*	-0.013
se (CI)	0.0064	0.0190	0.0200
hypothesis testing (t-values)			
65-			-3.88

*: significant at 5%.