Luxembourg Income Study Working Paper No. 209

Income Inequality Decomposition by Income Source and by Population Subgroups: A Theoretical Overview and the Empirical Case of Denmark

Jesper Drescher

August 1999

University of Southern Denmark **August 1999**

Institute of Economics, Odense University

Seminar

Income inequality decomposition by income source and by population subgroups. A theoretical overview and the empirical case of Denmark.*

by

Jesper Drescher

Skovgyden 26

DK-5230 Odense M

Denmark

E-mail: jesper.drescher@teliamail.dk

Telephone: 66178270

---

*The first version of this paper was prepared in April 1999 in Danish as a Bachelor seminar. This version, August 1999 has been shortened but the first version is availible on request, from the author. Mainly the theoretical sections have been shortened. The author do appreciate comments on this paper.

CONTENTS

1.Introduction

Analyses of income inequality are often applied to the case of measuring total income inequality in a given population, but often it is more interesting to make research into how income inequality can be related to income differerences between population subgroups or how income inequality can be related to the distribution of different income sources. The first category includes the cases, where total inequality can be expressed as the sum of a within-group inequality term and a between-group term, where the within-group contribution is itself a weighted sum of the sub-group inequality values. The second category includes the cases, where the contributions of separate income sources to total income inequality are examined. Furthermore combinations of the two categories can be done.

Many empirical studies have decomposed total income inequality by population subgroups or by income source, without being critical of which inequality measures that are able to be used - satisfactorily - in a decomposition. Therefore we begin in section 2 with a number of general principels that we might wish to be satisfied by any decomposition by population subgroups. Section 3 proposes some general principels that we might prefer to be satisfied by any decomposition by income sources. Section 4 gives a short overview of the data used in the following empirical studies. The empirical results of this paper are calculated on the basis of Luxembourg Income Study(LIS).[1] Section 5 and section 6 present the empirical work of this paper. Finally section 7 summarizes the main findings of this study.

2.Decomposition by population subgroups: Theory

When analyzing income inequality within and between subgroups the population is split up into K disjoint subgroups G_1,........, G_K where subgroup k consist of $N_k(\geq 1)$ individuals. The total population consist of N individuals and because the subgroups are disjoint $N = \sum_{k=1}^{K} N_k$

[1]Technically, the LIS-database is used via e-mail. The researcher set up some jobs using SPSS or SAS software, these jobs are sent via e-mail to a server address in Luxembourg. The jobs are automatically executed and the output file is sent back to an e-mail address specified by the user. Further information is available on the internet: **http://lissy.ceps.lu**

Mean income of subgroup k, μ_k, is

$$\mu_k = \frac{\sum_{j \in G_k} x_j}{N_k} \quad \text{, where } x_j = \text{individual j´s income} \qquad (2.1)$$

Shorrocks(1984,1370) shows that the condition for additive decomposability may be written

$$I(\mathbf{x}) = I(\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_K) = \sum_{i=1}^{K} w_k I(\mathbf{x}_k) + B \qquad (2.2)$$

where $\mathbf{x}_1,.....\mathbf{x}_K$ represents any partition of the distribution **x** into K subgroups, and where the coefficients w_k and the between-group term, B, depend only on $\boldsymbol{\mu}$ and **N**, where $\boldsymbol{\mu}=(\mu_1,........,\mu_K)$ and $\mathbf{N}=(N_1,........, N_K)$. B is assumed to be independent of inequality within the individual subgroups. Making within-group transfers until $x_j^k=\mu_k$ in each subgroup and letting $\boldsymbol{\mu}_{Nk}$ represent the unit vector with N_k components, we obtain

$$I(\mu_1 \boldsymbol{\mu}_{N1}, \mu_2 \boldsymbol{\mu}_{N2},........, \mu_K \boldsymbol{\mu}_{NK}; N)=B \qquad (2.3)$$

by assuming that 1) I(**x**) obtains it´s minimum 0 when $x_j=x_i$, for i=1,....,N and j=1,....,N and 2) B is invariant to transfers within the subgroups.

Given the above assumptions an additive decomposition of the inequality measure I satisfies the constraint

$$I(\mathbf{x}) = I(\mathbf{x}_1, \mathbf{x}_2, ...,\mathbf{x}_K) = \sum_{k=1}^{K} w_k I(\mathbf{x}_k) + I(\mu_1 \boldsymbol{\mu}_{N1}, \mu_2 \boldsymbol{\mu}_{N2},........, \mu_K \boldsymbol{\mu}_{NK}; N) \qquad (2.4)$$

Shorrocks(1980,615-619) shows that an inequality measure which is said to be additively decomposable and furthermore differentiable, must take the form

$$I(\mathbf{x}) = \frac{1}{\theta(\mu, N)} \sum_i \{\phi(x_i) - \phi(\mu)\} \qquad (2.5)$$

where μ and N respectively denote the mean and population size of **x** and $\theta(\mu,N)$ is positive;

$d\theta(\mu,N)/d\mu$ and $d\varphi(\mu)/d\mu$ are continuous; and $\varphi(\cdot)$ is strictly convex.

If we, in addition, assume two further proporties that we should like I(**x**) to satisfy:

1) I(**x**)is homogeneous of degree zero in **x**, which implies income homogeneity
2) I(**x**,**x**,.........**x** ; rN) = I(**x** ; N) for any positive integer r. Here we introduce the principle of population replication. The principle states that if r groups, each containing N individuals and having an identical distribution **x**, are aggregated into a single population of rN individuals, then aggregate inequality is the same as in each of the constituent groups

Shorrocks(1980, 617-625) has shown that the additively decomposable indices satisfying both income homogeneity and population replication consist of the "Generalized Entropy" family [2]. The "Generalized Entropy" family - for the total population - is by definition

$$I_c(\mathbf{x}) = \frac{1}{Nc(c-1)}\sum_{i=1}^{N}\left[\left(\frac{x_i}{\mu}\right)-1\right], c \neq 0,1 \qquad (2.6)$$

$$I_0(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{\mu}{x_i}\right), c = 0 \qquad (2.7)$$

$$I_1(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{x_i}{\mu}\right)\log\left(\frac{x_i}{\mu}\right)\right], c = 1 \qquad (2.8)$$

Shorrocks(1980, 621) shows that the "Generalized Entropy" family can be decomposed by population subgroups and defined as following

[2] The Gini coefficient is known to be additively decomposable when the incomes in one subgroup are all less than those in the other subgroup(Pyatt, 1976).

$$I_c(\mathbf{x}) = \sum_{k=1}^{K} \frac{N_k}{N}\left(\frac{\mu_k}{\mu}\right)^c \cdot I_{c,k} + \sum_{k=1}^{K} \frac{N_k}{N} \frac{1}{c(c-1)}\left(\left(\frac{\mu_k}{\mu}\right)^c - 1\right) \tag{2.9}$$

where: $I_{c,k}$ is the entropy measure for $c \neq 0$ and1 in group k.

$$I_0(\mathbf{x}) = \sum_{k=1}^{K} \frac{N_k}{N} \cdot I_{0,k} + \sum_{k=1}^{K} \frac{N_k}{N} \cdot \log\left(\frac{\mu}{\mu_k}\right) \tag{2.10}$$

where: $I_{0,k}$ is the entropy measure for c=0 in group k.

$$I_1(\mathbf{x}) = \sum_{k=1}^{K} \frac{N_k \cdot \mu_k}{N \cdot \mu} \cdot I_{1,k} + \sum_{k=1}^{K} \frac{N_k \cdot \mu_k}{N \cdot \mu} \log\left(\frac{\mu_k}{\mu}\right) \tag{2.11}$$

where: $I_{1,k}$ is the entropy measure for c=1 in group k.

The decomposition coefficients for these indices are given by

$$w_k = \frac{N_k}{N}\left(\frac{\mu_k}{\mu}\right)^c \tag{2.12}$$

and sum to unity only when c=0 or c=1. Thus, in general, the total within-group contribution to inequality $\sum w_k I(\mathbf{x}_k)$ will not be a weighted average of the subgroup values $I(\mathbf{x}_k)$. Theil(1967, 125) has pointed out a serious implication about that. It can be shown that $1-\sum w_k$ is proportional to B i (2.2). Thus, apart from the two measures proposed by Theil (c=0 and c=1)[3], the decomposition coefficients are not independent of the between group contribution.

Only when w_k is independent of μ_k, total inequality can - unambiguously - be split into the contribution due to differences between subgroups

$$B = \frac{1}{N}\sum_{k=1}^{K} N_k \log\left(\frac{\mu}{\mu_k}\right), \tag{2.13}$$

[3]In the literature on income distribution c=1 is known as Theils coefficient and c=0 as the mean log deviation(Sørensen 1999, Ch. 3, p.5).

and the contribution due to inequality within each subgroup k=1,.....K,

$$C_k = w_k I(\mathbf{x}_k) = \frac{1}{N}\sum_{j=1}^{N_k}\log\left(\frac{\mu}{x_j^k}\right), \tag{2.14}$$

in such a way that total inequality is the sum of these K+1 contributions.

3.Decomposition by income source: Theory

This section disaggregates the income of individuals or households into different factor components, such as wages, interest, transfer payments etc. and considers how to assess the contributions of these income sources to total income inequality.

Shorrocks(1982, 196-205) has shown that on the assumptions of

1) K disjoint and exhaustive components of income can be identified, which implies

$$x_i = \sum_{k=1}^{K} x_i^k, i = 1,\ldots\ldots, N \tag{3.1}$$

where: x_i=total income of individual i,

x_i^k=income of individual i from source k.

$\mathbf{X}$=the distribution of total income

$\mathbf{X}^k$=the distribution of income from source k

the contribution of factor k to total income inequality can be represented by

$S_k(\mathbf{X}^1,\ldots\ldots,\mathbf{X}^K;K)$

2) $I(\mathbf{X}^k)=0$ if and only if $\mathbf{X}^k=\mu^k\mathbf{e}$, where μ^k=mean of income source k and $\mathbf{e}=(1,1,\ldots\ldots..1)$

3) $S_k(\mathbf{X}^1,\ldots\ldots,\mathbf{X}^K;K)$ is continuous in $\mathbf{X}^k$ and income factors are treated symetrically.

$S_k(\mathbf{X}^1,\ldots\ldots,\mathbf{X}^K;K) = S_{\prod k}(\mathbf{X}^{\prod 1},\ldots\ldots,\mathbf{X}^{\prod K};K)$, where $\prod 1,\ldots\ldots.\prod K$ is any permutation of 1,........,K

4) $S_1(\mathbf{X}^1,......,\mathbf{X}^K;K)=S_1(\mathbf{X}^1, \mathbf{X}-\mathbf{X}^1;2)=S(\mathbf{X}^1, \mathbf{X})$. The assumption implies that the contribution of any one factor does not depend on the level of disaggregation. Otherwise the contribution of transfer payments might change if they were split into pensions, disability pay, unemployment pay, and so on. Because of assumption 3), which implies that the factors are treated symmetrically we have $S_k(\mathbf{X}^1,......,\mathbf{X}^K;K)=S(\mathbf{X}^k, \mathbf{X})$

5) $$\sum_{k=1}^{K} S_k\left(\mathbf{X}^1,......,\mathbf{X}^K;K\right) = \sum_{k=1}^{K} S\left(\mathbf{X}^k,\mathbf{X}\right) = I(\mathbf{X}) \tag{3.2}$$

this last assumption requires that the decomposition is consistent, in the sense that the contributions sum to the overall amount of inequality

a decomposition of inequality indices can be written as

$$S\left(\mathbf{X}^k,\mathbf{X}\right) = \mathbf{a}(\mathbf{X})\cdot\mathbf{X}^k = \sum_{i=1}^{N} a_i(\mathbf{X})x_i^k \tag{3.3}$$

where

$$I(\mathbf{X}) = \mathbf{a}(\mathbf{X})\cdot\mathbf{X} = \sum_{i=1}^{N} a_i(\mathbf{X})x_i.$$

The proportional contribution of factor k when inequality is measured by I is given in (3.4)

$$s_k = \frac{S_k}{I(\mathbf{X})} = \frac{\mathbf{a}(\mathbf{X})\cdot\mathbf{X}^k}{\mathbf{a}(\mathbf{X})\cdot\mathbf{X}} = \frac{\sum_{i=1}^{N} a_i(\mathbf{X})x_i^k}{\sum_{i=1}^{N} a_i(\mathbf{X})x_i} \tag{3.4}$$

(3.3) suggest that when an inequality index is written as a weighted sum of incomes, the decomposition contribution of factor k is the same weighted sum applied to factor k. The in

(3.3) given "natural" decomposition rule implies the following decompositions of a number of indices

In the Gini index, $a_i(\mathbf{X})$ is given by

$$a_i(\mathbf{X}) = 2\left(i - \frac{N+1}{2}\right) / N^2 \cdot \mu \tag{3.5}$$

from (3.3) and (3.5) we obtain

$$S(\mathbf{X}^k, \mathbf{X}) = \frac{2}{N^2 \mu} \sum_{i=1}^{N} \left(i - \frac{N+1}{2}\right) x_i^k = \frac{\mu_k}{\mu} \bar{G}(\mathbf{X}^k) \tag{3.6}$$

where μ_k=mean income of income source k.
$\bar{G}(\mathbf{X}^k)$ = Pseudo-Gini coefficient of factor k[4]

We can write the Gini index as

$$G(\mathbf{X}) = \sum_{k=1}^{K} S(\mathbf{X}^k, \mathbf{X}) = \sum_{k=1}^{K} \frac{\mu^k}{\mu} \bar{G}(\mathbf{X}^k) \tag{3.7}$$

The proportional contribution of factor k when inequality is measured by the Gini index is given by

$$s_k(G) = \frac{\sum_{i=1}^{N} \left(i - \frac{N+1}{2}\right) x_i^k}{\sum_{i=1}^{N} \left(i - \frac{N+1}{2}\right) x_i} \tag{3.8}$$

The factor decomposition of the Theil coefficient suggest using the weights

$$a_i(\mathbf{X}) = \frac{1}{N\mu} \log\left(\frac{x_i}{\mu}\right) \quad , \tag{3.9}$$

and the Theil value gives the factor contributions

$$S(\mathbf{X}^k, \mathbf{X}) = \frac{1}{N\mu} \sum_{i=1}^{N} \log\left(\frac{x_i}{\mu}\right) x_i^k \tag{3.10}$$

---

[4]It is not the conventional Gini value since weights attached to x^k_i correspond to the rank of individual i in the distribution of **X**; $x_1 \leq x_2 \leq ....\leq x_N$, which is not in general the same as his rank in the distribution of $\mathbf{X}^k$(Sørensen 1999).

The Theil index can be written as

$$I_1(\mathbf{X}) = \sum_{k=1}^{K} S\left(\mathbf{X}^k, \mathbf{X}\right) = \sum_{k=1}^{K} \frac{1}{N\mu} \sum_{i=1}^{N} \log\left(\frac{x_i}{\mu}\right) x_i^k \tag{3.11}$$

The proportional contribution of factor k when inequality is measured by $I_1(\mathbf{X})$ is given by

$$s_k\left(I_1\right) = \frac{\sum_{i=1}^{N} x_i^k \log\left(\frac{x_i}{\mu}\right)}{\sum_{i=1}^{N} x_i \log\left(\frac{x_i}{\mu}\right)} \tag{3.12}$$

Similar results are obtained when $I_0(\mathbf{X})$ is decomposed

$$I_0(\mathbf{X}) = \sum_{k=1}^{K} \frac{1}{N} \sum_{i=1}^{N} \log\left(\frac{\mu}{x_i}\right) \cdot \frac{1}{x_i} \cdot x_i^k \ , \tag{3.13}$$

$$s_k\left(I_0\right) = \frac{\sum_{i=1}^{N} \frac{x_i^k}{x_i} \log\left(\frac{\mu}{x_i}\right)}{\sum_{i=1}^{N} \log\left(\frac{\mu}{x_i}\right)} \tag{3.14}$$

and when $I_c(\mathbf{X})$ is decomposed

$$I_c(\mathbf{X}) = \sum_{k=1}^{K} \frac{1}{Nc(c-1)} \sum_{i=1}^{N} \left[\left(\frac{x_i}{\mu}\right)^c - 1\right] \cdot \frac{1}{x_i} \cdot x_i^k \, , \tag{3.15}$$

$$s_k\left(I_c\right) = \frac{\sum_{i=1}^{N} \frac{x_i^k}{x_i} \left[\left(\frac{x_i}{\mu}\right)^c - 1\right]}{\sum_{i=1}^{N} \left[\left(\frac{x_i}{\mu}\right)^c - 1\right]} \tag{3.16}$$

The above decompositions of inequality indices are given as “natural” decomposition rules, but there is no reason why they should be given special attention because (3.3) does not give the coefficients, $a_i(\mathbf{X})$, unambiguously. The coefficients, which meet the assumptions of (3.3), are an infinite number , because (3.3) only places one restriction on N coefficients. Shorrocks(1982, 199-202) has shown that there is an infinite number of decompositions by income source and the contribution of any factor expressed as a proportion of total inequality, $s_k(\circ)$, can be carried out to give any value in the interval $(-\infty, \infty)$. Shorrocks suggests two assumptions, which implies an unambiguous decomposition

6) Population symmetry: If $\mathbf{P}$ is any N*N permutation matrix, $S(\mathbf{X}^k\mathbf{P}, \mathbf{XP}) = S(\mathbf{X}^k, \mathbf{X})$

7) Two factor symmetry: If $\mathbf{P}$ is any permutation matrix,
$S(\mathbf{X}^1, \mathbf{X}^1+\mathbf{X}^1\mathbf{P}) = S(\mathbf{X}^1\mathbf{P}, \mathbf{X}^1+\mathbf{X}^1\mathbf{P})$

If assumption 6) is accepted , there is a unique decomposition rule for a two person population. The contribution proportions will be invariant to the choice of inequality measure and correspond to those obtained in the natural decomposition of the variance(or Gini, since the expressions are identical with only two individuals)[5]. With $N \geq 3$ we do not have any unique decomposition rule. $N \geq 3$ is relevant in any empirical case and therefore assumption 7) is introduced and the contribution of any factor expressed as a proportion of total inequality, $s_k(\circ)$, is given by

$$s_k(I) = \frac{S(\mathbf{X}^k, \mathbf{X})}{I(\mathbf{X})} = \frac{\operatorname{cov}(\mathbf{X}^k, \mathbf{X})}{\sigma^2(\mathbf{X})} \quad \text{for all } \mathbf{X} \neq \mu\mathbf{e} \tag{3.17}$$

(3.17) has two important consequences. First assumption 7) has eliminated all the degrees of freedom associated with the choice of $a_i(\mathbf{X})$.That gives us a unique decomposition rule for any inequality measure. Second (3.17) gives the relative importance of different income components, independent of the choice of inequality measure[6]. Furthermore (3.17) corresponds to the “natural” decomposition - by income source - of the square of the

[5]The decomposition will be unique when N=2, because assumption 6) eliminates one degree of freedom.

[6]This proporty would not hold if we insisted that only natural decomposition rules were used, since a change in inequality measure would simultaneously change the decomposition rule.

coefficient of variation.

The coefficient of variation is written as

$$CV = \frac{\sigma^2(\mathbf{X})}{\mu^2} \qquad (3.18)$$

A problematic issue - which is ignored above - concerns the links between different kinds of income. The rationale underlying any factor decomposition requires that we examine each income component separately and neglect the feedback effects on other income sources. (Shorrocks 1982, 210). Furthermore a decomposition using post tax incomes takes no account of the impact of taxes on the distribution of pre-tax incomes.

4.Data

Since the foundation of Luxembourg Income Study(LIS) in 1983 many researchers have improved in understanding inequality and the importance of which inequality measures have been used. The LIS database includes about 80 datasets for 25 countries, covering the period 1967-1997[7]. Most datasets include three different files, the first with data at the household level, the second at the individual level and the third at the child level. In the dataset concerning Denmark the unit is consistently a D-family(see appendix 1 for the definition of a D-family) and not the household as in most other LIS datasets. The main issue in setting up such a database is to elaborate data from single national household income surveys and transforming variables and undo definitions in order to allow some satisfactory cross-country research. Perfect validity will never be reached but the LIS database allows a good degree of validity when cross-country income distribution research is done.

All datasets used in this paper concern Denmark. Datasets concerning Denmark in LIS database have been provided by official records created for administrative purposes. The datasets are based on income tax surveys and have been collected by the Danish Central

[7] Lisification process has started, and some datasets from 1997 will soon be included.

Statistical Office. The LISification process was done by the Danish Ministry of Economic Affairs and not by the LIS staff in Differdange. The sample for 1987 consists of 0.5% of all D-families in Denmark. This way, 12,516 D-families with 25,706 familymembers have been selected. The dataset of 1992 consists of 12,895 D-families(Kristensen, 1994). All families are given the weight 1 in the empirical work of this paper.

Comparative studies of income inequality are heavily affected by the "technical-parameters" chosen by the reaserchers. Most reaserchers agree that there are different economies of scale to households of different size and it seems reasonable to postulate the existence of positive economics of scale within larger households, but still there is no agreement on a "best" equivalence scale (Figini 1998, 4). It is outside the scope of this paper to make any conclusions about the choice of equivalence scale and the "OECD-scale" will be used in the following empirical work. The "OECD-scale" add the weight of 1 for the household head and the weights 0.7 and 0.5 for other adults and children (Atkinson, Rainwater, Smeeding 1995).[8] We assume that total income is shared equally among all the family members.
Figini (1998, 13) shows that inequality measures are heavily dependent on the type of equivalence scale used to adjust for household size and composition.
Head of household is defined as the family member with the highest income disregarding age and gender, i.e.[9]

Although Rawls (1971) has argued that, for ethical reasons, public policy should emphasize most the well-being of the least well-off, in practice reseachers face a problem in how to treat very low incomes. For example the 1987 Danish dataset contains an income of DKK.-656,959. In the following such observations are deleted, because they would otherwise dominate the

---

[8]The equivalent income is $$X = \frac{Y}{1 + 0,7(N_v - 1) + 0,5N_b}$$

Where Y=income of the family
N_v=number of adults in the family
N_b=number of children in the family.

[9]Please note the difference between the definition used by LIS, where the head of household is defined as the oldest male and the definition used by the Danish Ministry of Economic Affairs.

measures of inequality totally.[10] If we assume that negative incomes are selfemployed with large capital losses in a particular year, but who are otherwise well off, the inequality measures are more true after the redefinition. Furthermore some inequality indices are not defined for negative observations.

5.Decomposition by population subgroups: The empirical case of Denmark

The form of an unambiguous decomposition of income inequality by population subgroups was derived in section 2 and we ended up with a special case of the "Generalized Entropy" family where c=0(2.10). By insisting on a unique decomposition we do simultaneously accept the mean log deviation as a useful inequality measure. The decomposition by population subgroups below is using I_0 and therefore total inequality is the sum of K+1 contributions(see (2.13) and (2.14))

5.1 Decomposition by the educational attainment of the head of household

The decomposition procedure, which has been derived above has been applied to the Danish data and the families have been divided into seven groups on the basis of the educational attainment of the head of the family. Table 5.1.1 and table 5.1.2 show the decomposition of the mean log deviation in 1987 and 1992. The total income inequality in 1987 measured by $I_0(\mathbf{x})$ is 0.2952[11]. 0.1240 of the 0.2952 is related to inequality within the group "Level undefined". The relative large contribution should not surprise us, because the group must by definition be one of the most heterogeneous. The group "Lev1 Complete" does only contain 4 members and therefore its contribution to total inequality is insignificant. The contributions from inequality within the groups "Lev2. 1st stage" and "Lev2. 2nd stage" are 0.0623 and 0.0691. The contributions (within) from the last three groups "Lev3. 1st stage", "L3s2. univ degree" and "L3s3. univ degree" are 0.0068, 0.0119 and 0.0053. The sum of inequality within the seven groups is 0.2794. i.e. 94.65% of total income inequality can be related to income inequality within the seven groups. The remaining 5.35% can be related to income inequality between the seven groups.

---

[10] In the data programming the observations are not deleted but redefined, so that all values less than 0 are equal to 0.01

[11] Total income is defined as Disposable income.

Table 5.1.1. Decomposition by the educational attainment of the head of household in Denmark 1987.[a]

Level of education [b] k	Number of members of group k	Income inequality within group k	Contributin to total income inequality within group k	Mean income in group k[d]	Income inequality between the groups
Level undefined[c]	3724	0.4151	0.1240	70101	0.0813
Lev1 Complete	4	0.1443	0.0000	150498	-0.0002
Lev2. 1st stage	3096	0.2507	0.0623	95526	-0.0093
Lev2. 2nd stage	4075	0.2112	0.0691	100193	-0.0278
Lev3. 1st stage	424	0.2004	0.0068	107005	-0.0051
L3s2. univ degree	728	0.2042	0.0119	110012	-0.0104
L3s3. univ degree	411	0.1593	0.0053	135515	-0.0128
Total	12462	0.2952	0.2794	92028	0.0158

a) The used income inequality index is the mean log deviation.

b) 5 digit code according to ISCED(International Standard Classification of Education)see Kristensen(1994).

c) Incl. No answer.

d) Mean of the equavilent incomes.

Source: Own calculations on the basis of LIS database.

Table 5.1.2 shows that total income inequality in Demark - measured by $I_0(\mathbf{x})$ - has dropped to 0.2478 in 1992. The fall in inequality is mainly due to a fall in inequality within the seven groups and the sum of inequality within the seven groups has dropped to 0.2327. However inequality within the group “L3s3. univ degree” has risen from 0.1593 in 1987 to 0.1670 in 1992. The relative contribution to total inequality from inequality between the seven groups is in 1992 equal to 6.05%.

Table 5.1.2. Decomposition by the educational attainment of the head of household in

Denmark 1992.[a]

Level of education[b] k	Number of members of group k	Income inequality within group k	Contributin to total income inequality within group k	Mean income in group k[d]	Income inequality between the groups	
Level undefined[c]	3151	0.3714	0.0907	83792	0.0552	
Lev1 Complete	2	0.0361	0.0000	87058	0.0000	
Lev2. 1st stage	3361	0.2137	0.0557	97404	0.0197	
Lev2. 2nd stage	4496	0.1813	0.0632	110766	-0.0185	
Lev3. 1st stage	514	0.1074	0.0043	123077	-0.0063	
L3s2. univ degree	857	0.1828	0.0122	133457	-0.0159	
L3s3. univ degree	514	0.1670	0.0067	169803	-0.0191	
Total	12895	0.2478	0.2327	105040	0.0150	

a) + b) +c) +d) see table 5.1.1.

Source: Own calculations on the basis of LIS database.

5.2 Decomposition by working status of the head of household.

The Danish families are in table 5.2.1 and table 5.2.2 divided into groups on the basis of the working status of the head of household. The grouping results in a disaggregation of total income inequality into nine contributions - eight contributions from inequality within each of the eigth groups and one contribution from inequality between the eight groups. In 1987 0.1477 of total income inequality(50%) is due to inequality within the group “Selfemployed, farm”. Furthermore it is noteworthy that the mean log deviation within the group “Selfemployed, farm” is as high as 1.4157. If it is true in general that inequality within the group “Selfemployed, farm” is very large, it might be disirable to carry out standardized calculations when one is doing cross-country studies. The contributions from inequality within the groups “Selfempl.. nonfarm”, “Wage earners unskilled” and “Employees not further specified” are all less than 0.01, even though $I_{0,k}$ from (2.10) vary between 0.0836 and 0.2678.

Table 5.2.1. Decomposition by working status of the head of household in Denmark 1987.[a]

Type of worker k	Number of members of group k	Income inequality within group k	Contributin to total income inequality within group k	Mean income in group k[b]	Income inequality between the groups
Selfemployed, farm	1300	1.4157	0.1477	90140	0.0022
Selfempl.. nonfarm	230	0.2678	0.0049	67953	0.0056
Salaried employees	692	0.3697	0.0205	100579	-0.0049
Wage earners. skilled	3140	0.0870	0.0219	112143	-0.0498
Wage earners unskilled	1238	0.0836	0.0083	103531	-0.0117
Employees not further specified	1331	0.0873	0.0093	96878	-0.0055
Retired persons	1475	0.1338	0.0158	93110	-0.0014
Others	3056	0.1875	0.0460	64743	0.0862
Total	12462	0.2952	0.2745	92028	0.0207

a) The used income inequality index is the mean log deviation.

b) Mean of the equavilent incomes.

Source: Own calculations on the basis of LIS-database.

The sum of inequality within the eight groups is 0.2745. i.e. 92.99% of total income inequality can be related to income inequality within the eight groups. The remaining 7.01% can be related to income inequality between the seven groups.

If table 5.2.1 and table 5.2.2 are compared it can be seen that the composition of income inequality is on its way towards a more homogeneous income distribution within the population subgroups. On the other hand income inequality between the eight groups has risen from 1987 to 1992. In 1992 the contribution from inequality between the eight groups is 0.0233.

Table 5.2.2. Decomposition by working status of the head of household in Denmark 1992.[a]

Type of worker k	Number of members of group k	Income inequality within group k	Contributin to total income inequality within group k	Mean income in group k[b]	Income inequality between the groups
Selfemployed, farm	1475	1.0778	0.1233	69673	0.0470
Selfempl.. nonfarm	201	0.2294	0.0036	96258	0.0014
Salaried employees	721	0.2163	0.0121	126508	-0.0104
Wage earners. skilled	3932	0.0895	0.0273	129828	-0.0646
Wage earners unskilled	1263	0.0503	0.0049	116536	-0.0102
Employees not further specified	1655	0.0630	0.0081	104195	0.0010
Retired persons	383	0.1995	0.0059	93397	0.0035
Others	3265	0.1552	0.0393	84313	0.0557
Total	12895	0.2478	0.2245	105040	0.0233

a)+b) see table 5.2.1.

Source: Own calculations on the basis of LIS database.

The results, obtained from the two decompositions above, show that more than 90% of total income inequality can be related to income inequality within the groups. It is important to make it clear that the results are not true in general. Imagine that the population is divided into N subgroups and N_k=1 for k=1,2,....K. That would give us $I_{0,k}$=0 for k=1,......,K and the total income inequality can be related to inequality between the K groups.

6.Decomposition by income source: The empirical case of Denmark

Section 3 shows that the only unambiguously decomposition of income inequality by income source is given as (3.17), which corresponds to the "natural" decomposition of the square of the coefficient of variation. As noted in section 3, (3.17) gives the relative importance of different income components, independent of the choice of inequality measure. Table 6.1 shows the decomposition of income inequality by income source in Denmark 1987 corresponding to (3.17). Column 5 in table 6.1 indicates that mainly gross wages and salaries play a significant role in determining the overall income inequality. The contribution to total inequality is equal to 67.98%(see appendix 3 for calculations of total income inequality). The very significant role played by gross wages and salaries in determining the overall income inequality reflects, in fact, the impact of the share of this income source (64.81%) in total income rather than that of the inequality of the distribution of gross wages and salaries. Another income source playing a significant role is non-farm self-employment income. Its contribution to total income inequality is 12.37% and the significant contribution is mainly due to the inequality of the distribution of the income source, because the share of this income source of total income is equal to 3.27%. This fact should not surprise us, because only few families must be assumed to receive income from non-farm self-employment.[12] Furthermore two income sources play a significant role in determining the total inequality. Their contributions to total inequality are respectively equal to 12.74%(cash property income) and 12.61(means-tested social transfers). The positive significant role played by means-tested social transfers is remarkable when we consider the aim of means-tested social transfers, but the fact is that a mistake was made when the dataset was prepared. The payments in kind received by handicapped persons have been converted into cash benefits, which means that most handicapped persons have a very large income.[13]

Finally table 6.1 indicates that social transfers have a negative impact on total income inequality, which means that there is a negative correlation between the ranking of the source and that of total income. Although the “redistributing-effect” (-10.12%) of the source is relatively small compared to its share of total income(16.82%).

[12] The 1987 Danish dataset contains 1719 families, which received non-farm self-employment income.

[13] I owe this point to Professor Peder J. Pedersen, Institute of Economics, University of Aarhus.

Table 6.1. Decomposition by income source in Denmark 1987.

Income source: $\mathbf{X}^k$	Mean income of income source k, μ^k, in DKK.	The relative contribution to total income [a] of $\mathbf{X}^k$	Cov($\mathbf{X}^k$,$\mathbf{X}$)	The relative contribution to total income inequality of $\mathbf{X}^k$
Gross wages and salaries	85695	0.6481	4561084450	0.6798
Farm self-employment income	379	0.0029	87896612	0.0131
Non-farm self-employment income	4324	0.0327	829833028	0.1237
Cash property income	5698	0.0431	855074442	0.1274
Social transfers	22243	0.1682	-679108184	-0.1012
Means-tested social transfers	8665	0.0655	846220814	0.1261
Pensions	4443	0.0336	168230632	0.0251
Alimony or Child support	167	0.0013	12552538	0.0019
Other cash income	607	0.0046	27774206	0.0041
Total Gross Income	132220	1.0000	6709558538[b]	1.0000

a) Total income is defined as Total Gross Income.

b) Cov($\mathbf{X}$,$\mathbf{X}$)=σ^2($\mathbf{X}$).

Source: Own calculations on the basis of LIS database.

Table 6.2 shows the contributions of the income sources to total income inequality in 1992. It should be noted that means-tested social transfers has a negative impact on total income inequality in 1992, although it is almost insignificant(-0.0156). Furthermore, the redistribution-effect of social transfers has fallen to -7,53%, even though the share of social transfers in total income has risen to 21.13%. Finally the contribution of cash property income to total income inequality has risen to 25.21%.

Table 6.2. Decomposition by income source in Denmark 1992.

Income source: $\mathbf{X}^k$	Mean income of income source k, μ^k, in DKK.	The relative contribution to total income [a] of $\mathbf{X}^k$	Cov($\mathbf{X}^k$,$\mathbf{X}$)	The relative contribution to total income inequality of $\mathbf{X}^k$
Gross wages and salaries	96932	0,6254	7267123675	0,6400
Farm self-employment income	364	0,0023	179805183	0,0158
Non-farm self-employment income	4924	0,0318	1105671213	0,0974
Cash property income	6707	0,0433	2862741508	0,2521
Social transfers	32745	0,2113	-854819466	-0,0753
Means-tested social transfers	4954	0,0320	-176759959	-0,0156
Pensions	7074	0,0456	628804902	0,0554
Alimony or Child support	196	0,0013	40301785	0,0035
Other cash income	1103	0,0071	302772385	0,0267
Total Gross Income	154999	1,0000	11355641225[b]	1,0000

a) + b) see table 6.1.

Source: Own calculations on the basis of LIS database.

If we compare table 6.1 and table 6.2 to appendix 2 it can be seen that a decomposition by income source is sensitive to the choice of equivalence scale.

7.Conclusions

The number of inequality measures is large and the choice of inequality measure must be related to the aim of using it. This paper ends up with two extremes. If we insist on making an unambiguously decomposition by poulation subgroups we must accept the mean log deviation as inequality measure. On the other hand a decomposition by income source is - on the given assumptions - independent of the choice of inequality measure.

The empirical content of this paper is done on the basis of datasets concerning Denmark in LIS-database. A decomposition by the educational attainment of the head of household, where the mean log deviation is used as inequality measure shows that in 1987 94.65% of total income inequality can be related to inequality within the seven groups. In 1992 the share was 93.95%. Inequality between the groups contribute with the rest. A decomposition by working status of the head of household shows that about 50% of the total income inequality can be related to inequality within the group “selfemployed, farm” in 1987 and in 1992 as well. Furthermore about 7% of total inequality can be related to inequality between the eight groups in 1987 and in 1992 about 9.41% of total inequality can be related to inequality between the eight groups.

The empirical decomposition by income source shows that that gross wages and salaries have a significant impact on total income inequality. This very significant impact is mainly a consequence of the fact that the share of gross wages and salaries in total income is very large compared to other income sources. It is also proven that the income source “social transfers” has a negative impact on overall inequality which means that it is having a “redistributing-effect”. The “redistributing-effect” of social transfers is -0.1012% in 1987 and -0.0753 in 1992.

This paper has used datasets concerning the years 1987 and 1992. Section 5 and section 6 point out some differences between the composition of inequality in 1987 and the composition of inequality in 1992. One should, however, always be aware of uncertainty in the sample selection.

References

Atkinson A.B., Rainwater L., Smeeding T. 1995. *Income Distribution in OECD countries*. OECD, Paris.

Cowell, F. 1995. *Measuring Inequality.* Prentice Hall.

Figini, Paolo. 1998. *Inequality measures, equivalence scales and adjustment for household size and composition.* Working Paper No. 185. CEPS/INSTEAD, Differdange.

Flückiger Y.og Silber J. 1994. *Income inequality decomposition by income source and the breakdown of inequality differences between two population subgroups.* Working Paper No. 110. CEPS/INSTEAD, Differdange.

Kristensen, K. 1994. *Luxembourg Income Study - codebook and comments.* The Danish Ministry of Economic Affairs.

Pyatt, G. 1976. *On the Interpretation and Disaggregation of Gini Coefficients*. Economic Journal. 86, 243-255.

Rawls, J. 1971. *A Theory of Justice.* Harvard University Press, Cambridge, Massachusetts.

Sørensen, C. 1997. *Økonomisk Samfundsbeskrivelse, Demografi.* Systime.

Sørensen, C. 1999. *Økonomisk Samfundsbeskrivelse, Økonomisk fordeling.* Institute of Economics, Odense University, University of Southern Denmark.

Shorrocks, A.F. 1980. *The class of additively decomposable inequality measures.* Econometrica 48, 613-625.

Shorrocks, A.F. 1982. *Inequality decomposition by factor components*. Econometrica 50, 193-211.

Shorrocks, A.F. 1984. *Inequality decomposition by population subgroup*. Econometrica 52, 1369-1385.

Theil, H 1967. *Economics and Information Theory.* North Holland Publishing Company. Amsterdam.

Appendix 1: The unit

The unit in the sample is the household. A household has been defined as a “D-family”- an official concept at the Central Statistical Office.

A D-family is one of the following:

1) A married couple with og without children
2) An unmarried couple with common children. These children do not have to live together with their parents.
3) A single person with or without children
4) Two persons living at the same address, if
   - they do not have common children, and
   - they are of different sex, and
   - their age do not differ by more than 15 years, and
   - no other adults live at the same address
5) Two persons of the same sex officially registered as a couple
6) A child not living with its parents.

The above definition has been used since 1991. 1st of January 1991 the CSO changed the definition. Three important changes were made(Sørensen, 1997. 63-64):

- Children living together with their parents are now only a part of their parents family if they are under 18 years old. Before the change the limit was 26 years old.
- Two persons of the same sex officially registered as a couple are now one family.
- Two persons living at the same address are now one family if they fulfil the conditions of 4).

Appendix 2: Decomposition by income source without any use of equivalence scale.

Table b.2.1. Decomposition by income source in Denmark 1987.

Income source: $\mathbf{X}^k$	Mean income of income source k, μ^k, in DKK.	The relative contribution to total income [a] of $\mathbf{X}^k$	Cov($\mathbf{X}^k$,$\mathbf{X}$)	The relative contribution to total income inequality of $\mathbf{X}^k$
Gross wages and salaries	152856	0.6863	20340000000	0.7415
Farm self-employment income	573	0.0026	233070000	0.0085
Non-farm self-employment income	8569	0.0385	3401000000	0.1240
Cash property income	8456	0.0380	2384700000	0.0869
Social transfers	32518	0.1460	-1175000000	-0.0428
Means-tested social transfers	12574	0.0565	1946900000	0.0710
Pensions	5894	0.0265	145940000	0.0053
Alimony or Child support	265	0.0012	35060447	0.0013
Other cash income	1008	0.0045	118910000	0.0043
Total Gross Income	222711	1.0000	27430000000[b]	1.0000

a) Total income is defined as Total Gross Income.

b) Cov($\mathbf{X}$,$\mathbf{X}$)=σ^2($\mathbf{X}$).

Source: Own calculations on the basis of LIS database.

Table b.2.2. Decomposition by income source in Denmark 1992.

Income source: $\mathbf{X}^k$	Mean income of income source k, μ^k, in DKK.	The relative contribution to total income [a] of $\mathbf{X}^k$	Cov($\mathbf{X}^k$,$\mathbf{X}$)	The relative contribution to total income inequality of $\mathbf{X}^k$
Gross wages and salaries	170837	0.6643	28350000000	0.7075
Farm self-employment income	533	0.0021	577820000	0.0144
Non-farm self-employment income	9939	0.0386	6298000000	0.1572
Cash property income	9379	0.0365	4824400000	0.1204
Social transfers	48040	0.1868	-1268000000	-0.0316
Means-tested social transfers	7013	0.0273	-413200000	-0.0103
Pensions	9389	0.0365	805150000	0.0201
Alimony or Child support	259	0.0010	45747075	0.0011
Other cash income	1792	0.0070	856800000	0.0214
Total Gross Income	257181	1.0000	40072833124[b]	1.0000

a) Total income is defined as Total Gross Income.

b) Cov($\mathbf{X}$,$\mathbf{X}$)=σ^2($\mathbf{X}$).

Source: Own calculations on the basis of LIS database.

Appendix 3: Total income inequality in Denmark

Table b.3.1. Income inequality indices in Denmark. Income is defined as Total Gross Income.

Inequality index	1987	1992
Mean log deviation	0.3328	0.2967
Theil	0.1657	0.1692
Atkinson(ε=0.5)	0.0879	0.0860
Coefficient of variation	0.6195	0.6875
Gini	0.3079	0.3046

Note: Household inequality is measured using the "OECD-equivalence scale".

Source: Own calculations on the basis of LIS database.

Table b.3.2. Income inequality indices in Denmark. Income is defined as Disposable Income.

Inequality index	1987	1992
Mean log deviation	0.2952	0.2478
Theil	0.1358	0.1230
Atkinson(ε=0.5)	0.0719	0.0628
Coefficient of variation	0.5705	0.6347
Gini	0.2717	0.2465

Note: See table b.3.1.

Source: Own calculations on the basis of LIS database.