

THE TIME LAG IN ANNUAL HOUSEHOLD-BASED INCOME MEASURES: ASSESSING AND CORRECTING THE BIAS

Debels Annelies, Katholieke Universiteit Leuven

Vandecasteele Leen, University of Manchester

This is the peer reviewed version of the following article: Debels, A., & Vandecasteele, L. (2008). The Time Lag in Annual Household-Based Income Measures: Assessing and Correcting the Bias. Review of Income and Wealth, 54(1), 71-88., which has been published in final form at <https://doi.org/10.1111/j.1475-4991.2007.00260.x>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Annual income data are typically provided with a time lag. This article reviews several ways of dealing with this time lag in the construction of annual household-based income measures for individual economic well-being. It also proposes an alternative method that yields better estimates for equivalised household income, especially in the case of household composition change. Next, the two most commonly applied income measures are compared to this alternative measure with empirical income data from the European Community Household Panel. This comparison reveals that ignoring the time lag and household composition changes leads to substantial bias in income and poverty estimates and to erroneous conclusions about the determinants of poverty entry. The evidence in this article will be useful to researchers who want to make a well-informed choice between different annual income measures.

KEY WORDS: HOUSEHOLD INCOME / ANNUAL INCOME / POVERTY DYNAMICS / INCOME MEASUREMENT / EUROPEAN COMMUNITY HOUSEHOLD PANEL

THE TIME LAG IN ANNUAL HOUSEHOLD-BASED INCOME MEASURES: ASSESSING AND CORRECTING THE BIAS

1. Introduction

Measuring income through a survey is a foremost delicate enterprise, involving numerous decisions that may influence the outcome. One of the decisions to be taken concerns the reference period of income. Usually, surveys apply either an annual or a monthly reference period. An annual reference period is often preferred because it is less susceptible to short-term income fluctuations (Smeeding and Weinberg, 2001; Böheim and Jenkins, 2000). However, annual income measures entail an often-neglected problem: they are provided with a time lag because respondents can only report their annual income after the reference period has ended. This means that the income information provided in a survey usually refers to the (calendar) year prior to the survey.

Earlier research has already set out some of the problems caused by ignoring the time lag in annual income measures (Atkinson, Cantillon, Marlier and Nolan, 2002). A potential bias arises if researchers want to relate income to time-varying covariates or if their interests primarily lie in equivalised household income. In the latter situation, the household information on which the equivalised household income measure is based refers to the date of the survey, while income refers to the (calendar) year prior to this survey. So, as households change, there is a mismatch between the timing of income and the timing of household information (Cantillon, Van Dam, Van den Bosch and Van Hoorebeeck, 2003).

The present article will first compare existing annual income methods and evaluate how successful they are in dealing with the time lag. It will focus on household-based income

measures of individual economic well-being. Individual measures are often preferred in income research, especially in a longitudinal context (Walker, 1994; Jenkins, 2000).¹ Even with the individual as the unit of analysis, researchers usually choose a household-based measure because they want to take into account the resources and needs structure of the household in which individuals live.²

Additionally, an alternative income measure will be presented that not only corrects the time lag, but that also adjusts the equivalised household income for household composition changes during the measurement period. The general strategy in view of households with a changing composition is to register the income and needs of the household in which an individual happens to be at the time of the survey (e.g. Bane & Ellwood, 1986; Stevens, 1999; Whelan, Layte & Maître, 2003). This is not so much of a problem if the reference period of income falls shortly around the survey date. However, most annual income measures refer to the previous (calendar) year, and individuals may experience changing household contexts in the time frame of more than a year between the start of the reference period and the time of the survey. In this context, extrapolating the household resources and needs at the date of the survey to the income reference period might no longer be a valid assumption. The newly proposed income measure will adjust for the time lag in income measurement and take into account changes in household composition in the relevant period. It will be used as a benchmark against which the bias resulting from existing income methods can be assessed. We will compare the performance of the different income methods in varying types of analyses: cross-sectional income estimates, distributional indicators of income, poverty estimates and multivariate analyses of poverty dynamics.

In order to correct the time lag in income measures, longitudinal information is required. The data source chosen for the empirical part of this article is the European Community Household Panel (ECHP), a major source of income information for the countries of the EU-15.

2. Problems with annual household income measures

Assume that we have at our disposal a panel survey design with repeated survey interviews taken at the dates d , $d+1$, etcetera. These dates take place at a certain point in time in the respective calendar years T , $T+1$, etcetera. On each interview date, respondents report their annual income relating to a past period of one year duration. If the income measurement occurs at time d , then, depending on the survey questioning, the reported income may refer either to the twelve months prior to d or to the calendar year $T-1$ prior to d . We will take the latter survey questioning method as our starting point, because it reflects the situation in the dataset used further in the article. However, the problems and solutions set out in this article can be applied to both types of questioning. Assume now that we want to construct a household-based income measure for the economic well-being of an individual i during the calendar year T , drawing on the assumption that the income resources are pooled and equally shared within households.³

One common way to proceed is to follow what will be called the '*simple income method*' in this article. It consists of taking the sum of all personal resources within the household h_d to which the individual i belonged at the date of the interview d . This sum, further denoted as $Y_{h_d, T-1}$, is then divided by E_{h_d} , the equivalence scale of the household h in which the individual i resided at time d .⁴ The result of this computation is the equivalised household income Y_{hT}^* . This equivalised household income is subsequently attributed to each individual i in the household h at time d in order to obtain Y_{iT}^* :

$$(1) \quad Y_{iT}^* = \frac{Y_{h_d, T-1}}{E_{h_d}}$$

The simple income method is frequently encountered in the literature, among others in several reports from Eurostat (Dennis and Guio, 2004; Dennis and Guio, 2003; Mejer and Siermann, 2000). This method faces problems because the annual income is provided with a

time lag. In particular, it assumes that $Y_{h_d, T-1}$, the annual household income measured at date d, and thus referring to calendar year T-1, can be used as a proxy for $Y_{h_d, T}$, the household income during calendar year T. However, it has been widely demonstrated that income is not stable over time (Jenkins, 2000; Walker, 1994; Bane and Ellwood, 1986): inflation, economic ups and downs, policy changes, but also personal and household income dynamics are likely to lead to year-to-year changes in an individual's economic well-being. In addition, the simple income method creates a mismatch between the reference period of the household income and that of other time-varying covariates (Atkinson, Cantillon et al., 2002). This is especially problematic for researchers with an interest in the effect of certain life events on income or poverty dynamics, since the correct timing of income changes is crucial for explaining those dynamics.

These problems are easily resolved by acknowledging that $Y_{h_d, T-1}$, the annual household income measured at interview date d in year T refers to the previous calendar year T-1. Therefore, a second method that is often applied by researchers to get an estimate for the income in year T, is to lag the equivalised household income measured at date d+1 by one year (Fouarge and Layte, 2005; Layte and Whelan, 2003; Whelan, Layte and Maître, 2002; Whelan, Layte, Maître and Nolan, 2001):

$$(2) \quad Y_{i_h T}^* = \frac{Y_{h_{d+1}, T}}{E_{h_{d+1}}}$$

The method applied in equation 2 will be called the '*lagged income method*'. As is the case with the simple income method, the calculation of lagged equivalised income requires only one interview date. Yet, as soon as researchers want to relate the lagged income measure with time-varying covariates, more than one interview date will be needed⁵: one date d+1 for measuring equivalised income and one date d for measuring the corresponding time-varying covariate. In a panel data context, this would imply that for this type of analysis, the income measured in the first wave cannot be linked to a corresponding time-varying covariate, and this

wave of income information will be lost. For this reason, and especially when only a limited number of waves is available, researchers might prefer the simple over the lagged income method.

The lagged income method, however, also entails timing problems. While the household income $Y_{h_{d+1},T}$ refers to the calendar year T, the equivalence scale $E_{h_{d+1}}$ applies to the household composition at time d+1 in year T+1. In this way, a mismatch is due to arise between the reference period (year T) and the reference moment of the equivalence scale (d+1 in year T+1) (Cantillon, Van Dam et al., 2003). This is because household composition and size are subject to changes over time as children are born, household members die, relationships break up and new households are formed. Whenever an individual joined or left the household between the end of the calendar year T and interview date d+1, the lagged income strategy leads to incorrect equivalence scales (Cantillon, Van Dam et al., 2003).⁶ Take, for example, An who has been living together with David during the whole of 2006 and 2007 and who has given birth to a baby in 2007. A few months after the birth, An and David have been interviewed and have reported their income during the calendar year 2006, i.e. when the baby had not yet been born. When calculating this household's equivalised income for 2006, the household income of 2006 is divided by an equivalence scale referring to the new household composition in 2007. Because too high an equivalence scale is used, this results in an underestimation of the equivalised household income of 2006.⁷

An additional problem with the lagged income method arises whenever an income earner joined or left the household. The reason is that the household income $Y_{h_{d+1},T}$ is computed as the sum of the personal incomes of the members belonging to the individuals' household at date d+1. So, if an income earner has joined or left the household between the end of the calendar year T and d+1, the personal incomes are totaled over the wrong household members and a problem arises with the resulting sum of the personal incomes (Cantillon, Van Dam et al.,

2003). For example, if An had been living alone in 2006, and David had moved in with her before the date of the interview in 2007, then the personal income he earned in 2006 would have been added to the total household income for the calendar year 2006, even though he did not yet belong to this household in 2006.

Apart from the problems that arise because annual income is provided with a time lag, the simple and lagged methods suffer from a more fundamental stock-flow problem. Whenever annual income, a flow measure referring to a one-year period, is connected to household composition, a stock measure referring to a certain point in time, problems are due to arise because household composition can change throughout the period under consideration. Table 1 illustrates the importance of household composition changes in different countries. During one calendar year, on average 12.4 percent of the respondents experience one or more changes in household composition. This percentage varies between 9.4 percent in Greece and 15.8 percent in Ireland. The majority of the household changes occur within the same household setting, while a slight minority moves to a different household or splits off to form a separate household (between 0.8 percent and 2.6 percent). Changes within the same household most commonly involve changes in the number of adults (between 5.7 percent and 10.4 percent). Somewhat less frequent are changes in the number of children (between 1.9 percent and 4.3 percent), followed by changes in both the number of adults and the number children in the same year (between 0.3 percent and 1.1 percent).

Table 1. Percentage of respondents with household composition changes during one calendar year, by country, average over calendar years 1995-2000

| Country | Same household, changing composition: | | | | |
|-----------------|---------------------------------------|-----------------------------|---------------------------|--|-------------------------------|
| | No changes | Changing number of children | Changing number of adults | Changing number of children and adults | Moving to different household |
| Austria | 88.4 | 2.7 | 6.7 | 1 | 1.1 |
| Belgium | 88.7 | 2.5 | 7 | 0.4 | 1.5 |
| Denmark | 84.5 | 4.3 | 8.1 | 0.5 | 2.6 |
| Finland | 88.2 | 3.2 | 5.7 | 0.6 | 2.2 |
| Greece | 90.6 | 2 | 6.2 | 0.4 | 0.8 |
| France | 87.5 | 3.1 | 7.6 | 0.4 | 1.4 |
| Ireland | 84.2 | 3.5 | 10.4 | 0.9 | 1 |
| Italy | 89.9 | 2 | 6.6 | 0.6 | 1 |
| Luxembourg | 87.8 | 3.4 | 6.9 | 0.7 | 1.2 |
| The Netherlands | 88.8 | 3.3 | 6.7 | 0.3 | 1 |
| Portugal | 87.3 | 2.4 | 8.2 | 1.1 | 1 |
| Spain | 86.8 | 1.9 | 9.4 | 0.8 | 1.2 |
| United Kingdom | 86 | 2.7 | 8.5 | 0.7 | 2 |

Source: ECHP-UDB, 2003

One appropriate way to address this problem and at the same time the problems created by the time lag, is to measure household income for the subsequent households to which an individual i belonged throughout the year T , dividing each of these household incomes by the proper equivalence scale. We propose the following measure, which will be further denoted as the '*change-adjusted income method*':⁸

$$(3) \quad Y_{i_h T}^* = \sum_{h=1}^H t_h \cdot \frac{Y_{h_t}}{E_{t_h}} \quad \text{with} \quad \sum_{h=1}^H t_h = 1$$

where H is the total number of households h to which the individual belonged throughout the year T ; t_h denotes the duration of the period (measured as a proportion of the calendar year) in which the individual belonged to household h ; Y_{h_t} is the household income counted over the members belonging to household h during the period t_h and E_{t_h} is the equivalence scale corresponding to the household composition of the household h during the period t_h . The annual equivalised household income in year T is thus measured as the weighted sum of the equivalised household incomes in year T of the households to which the individual belonged. The weight applied to the incomes corresponds to the duration of the period that the individual resided in each household.

The change-adjusted income method is capable of adequately resolving the timing problems otherwise inherent to annual equivalised household income measures, but requires additional information about the occurrence and timing of household composition changes throughout calendar year T . Usually however, panel data do contain information about events such as births, deaths and movements of individuals into or out of the household. It also requires that all individuals are followed through time.

3. Data and operationalization

The ECHP is a harmonised cross-national household panel dataset focusing on personal and household living conditions. With its extensive coverage of income components, it has been widely used in income and poverty dynamics research (see among others: (Apospori and Millar, 2003; Cabrero, 2003; Kuchler and Goebel, 2003; Layte and Whelan, 2003; Whelan, Layte and

Maitre, 2003; Barnes, Heady and Middleton, 2002). The ECHP ran from 1994 until 2001 and comprises panel data for 14 member states of the European Union. In the empirical part of this study, we have included all countries, except Germany. The reason is that Germany did not provide information on the timing of household changes, so that the change-adjusted income measure could not be computed. The analyses for the United Kingdom and Luxembourg are based on respectively the British Household Panel Study (BHPS) and the Luxembourg Household Panel (PSELL), as integrated in the ECHP-data. All analyses are restricted to the initial sample of individuals, because these are the only persons who were longitudinally followed in the ECHP even if they moved to a different household. Indeed, for the change-adjusted income measure, information from two subsequent waves is needed. The analyses in this article refer to the time frame 1995-2000. The first year of the panel (1994) has been omitted because no accurate information on household composition changes was available⁹ and the last year of the panel (2001) has been left out because the income data referring to 2001 was not collected.

The simple and lagged income measures have been constructed according to the formulas 1-2 (see section 2) on the basis of the household income variable and the modified OECD-equivalence scale provided by Eurostat.¹⁰ The standard Eurostat correction for within-household nonresponse (see: Eurostat, 2002) has been removed from the ECHP household income variables used to calculate simple and lagged income, in order to warrant full comparability with the change-adjusted income measure. The change-adjusted income measure has been constructed by means of retrospective information about who entered and left the household in the months between the measurement dates d and $d+1$. For each month t in the calendar year T , we have determined the modified OECD-equivalence scale E_{h_t} corresponding to the composition of the household to which the individual belonged during this month. However, we have not been able to account for individuals present in a household for

short periods in between the interview dates $d, d+1, \dots$, because no information was recorded on these individuals. The household income Y_{h_t} during month t has been calculated as the sum of the annual personal incomes (reported at $d+1$) of the individuals belonging to the household in that month, complemented with the household components (see note 4) and divided by twelve.¹¹ Finally, to obtain an annual estimate, the sum is taken of all monthly equivalised incomes, according to the following equation:

$$(4) \quad Y_{i_h T}^* = \sum_{t=1}^{12} \frac{I}{12} \cdot \frac{Y_{h_t}}{E_{h_t}}$$

The reliability of this measure depends on the quality of the retrospective information about the household composition changes. Unfortunately, the information needed to retrospectively reconstruct household changes in the ECHP (i.e. year and month of the household interview, year and month of the move into/out of the household) is not complete. Over the waves 1995-2001, on average 0.34 percent of the individuals have missing information on the month of the interview and on average 0.04 percent have missing information on the year of the interview. In those cases, the year and month of the interview have been imputed by the modus of the year and the month of the interview for the country and wave in question. Some individuals of whom we know that they moved into or out of a household between two waves have missing information on the exact year and month of their move (on average 5.44% over the waves 1995-2001). In those cases, the month of the move has been imputed by the month in the middle between last year's and this year's interview date. Personal income amounts have not been imputed. For individuals with missing information for less than six out of the twelve months during one year, the average income per month has been calculated on the basis of the available months and has been multiplied by twelve. This has been done for on average 1.77% of the incomes over the waves 1995-2000.

Finally, it must be noted that all income amounts are net amounts, expressed in purchasing power parities (PPP's, see: Eurostat, 2003) to allow for comparisons between countries. They have been deflated by means of the country's Harmonised Index of Consumer Prices (HICP, European Central Bank, 2007) to allow for comparisons over years. Income poverty has been operationalised as 60 percent of the median equivalised household income for a given calendar year and country.

4. Descriptive comparison of the three measures

A first way of comparing the simple, lagged and change-adjusted income measure is to investigate the (dis)similarities in their distributions. This is achieved in Table 2 by comparing medians between the measures and by providing a test-statistic for differences between the income distributions.¹² The change-adjusted income distribution is taken as the basis for comparison. The p-values in Table 2 indicate the significance of the paired-difference Wilcoxon Signed Rank Tests (WSR) for differences between the respective distributions.¹³

Despite minor differences across countries and years, some general tendencies can be observed in Table 2. As could be expected, the simple income distribution deviates most from the change-adjusted one: the median of the simple income is consistently and substantially smaller than the median of change-adjusted income and the WSR-tests all point to highly significant differences between the distributions. Given that the simple income method takes the income of the previous calendar year as a proxy for the current year's income, this discrepancy is probably due to yearly rises in income that are not accounted for by deflating the income data. On the other hand, the smaller deviations between the change-adjusted and the lagged income distribution can be explained by the fact that the lagged income is identical to the change-

adjusted income whenever there was no change in household composition between the start of calendar year T and the measurement of the income at the interview date d+1.

Table 2. Distributional differences between the changed-adjusted and the simple and lagged income method (ECHP, 1995-1997-2000, longitudinally weighted)

| | | Change-adjusted income | Simple income | | Lagged income | |
|-------------------------------|----------------|------------------------|---------------|-------------|---------------|-------------|
| | | Median | Median | p-value WSR | Median | p-value WSR |
| Austria | 1997 (N=5643) | 13513 | 13366 | *** | 13461 | *** |
| | 2000 (N=4752) | 15644 | 14998 | *** | 15271 | *** |
| Belgium | 1995 (N=5183) | 13284 | 12955 | *** | 13140 | *** |
| | 1997 (N=4333) | 14349 | 13863 | *** | 14152 | *** |
| | 2000 (N=3332) | 14642 | 13913 | *** | 14550 | *** |
| Denmark | 1995 (N=4204) | 13419 | 13062 | *** | 13201 | *** |
| | 1997 (N=3222) | 14694 | 13908 | *** | 14407 | *** |
| | 2000 (N=2815) | 15769 | 15257 | *** | 15443 | *** |
| Finland | 1997 (N=6161) | 8145 | 7800 | *** | 8190 | *** |
| | 2000 (N=2289) | 8496 | 8080 | *** | 8379 | *** |
| France | 1995 (N=11467) | 11912 | 11308 | *** | 11720 | *** |
| | 1997 (N=9677) | 12510 | 12038 | *** | 12326 | *** |
| | 2000 (N=3936) | 13457 | 12768 | *** | 13208 | *** |
| Greece | 1995 (N=10284) | 7132 | 6323 | *** | 7064 | *** |
| | 1997 (N=8394) | 6775 | 6259 | *** | 6713 | *** |
| | 2000 (N=8135) | 7284 | 6916 | *** | 7278 | *** |
| Ireland | 1995 (N=5555) | 9938 | 9226 | *** | 9787 | *** |
| | 1997 (N=4805) | 11091 | 9847 | *** | 10954 | *** |
| | 2000 (N=3106) | 12003 | 10311 | *** | 11836 | *** |
| Italy | 1995 (N=15668) | 9647 | 8950 | *** | 9486 | *** |
| | 1997 (N=13487) | 9909 | 9135 | *** | 9786 | *** |
| | 2000 (N=11259) | 10763 | 10120 | *** | 10668 | *** |
| Luxembourg | 1997 (N=4713) | 20835 | 19682 | *** | 20751 | *** |
| | 2000 (N=3855) | 22854 | 21405 | *** | 22826 | *** |
| The Netherlands | 1995 (N=7455) | 11777 | 11357 | *** | 11662 | *** |
| | 1997 (N=6787) | 13256 | 12641 | *** | 13125 | *** |
| | 2000 (N=6033) | 13604 | 12850 | *** | 13554 | *** |
| Portugal | 1995 (N=10082) | 6582 | 6051 | *** | 6530 | *** |
| | 1997 (N=9022) | 6989 | 6504 | *** | 6876 | *** |
| | 2000 (N=8588) | 7916 | 7213 | *** | 7843 | *** |
| Spain | 1995 (N=12758) | 7690 | 7404 | *** | 7650 | *** |
| | 1997 (N=11369) | 8109 | 7478 | *** | 8014 | *** |
| | 2000 (N=9386) | 9880 | 8765 | *** | 9700 | *** |
| United Kingdom ⁽¹⁾ | 1997 (N=6385) | 13401 | 12807 | *** | 13225 | *** |
| | 2000 (N=5923) | 14576 | 13556 | *** | 14410 | *** |

Table 2. Continued

Note: WSR= paired-difference Wilcoxon Signed Rank test for differences in distribution between change-adjusted and simple/lagged income distribution

(1): The 1995 estimates for the United Kingdom could not be calculated because the HICP is not available.

*** = $p < 0.001$ ** = $p < 0.01$ * = $p < 0.05$

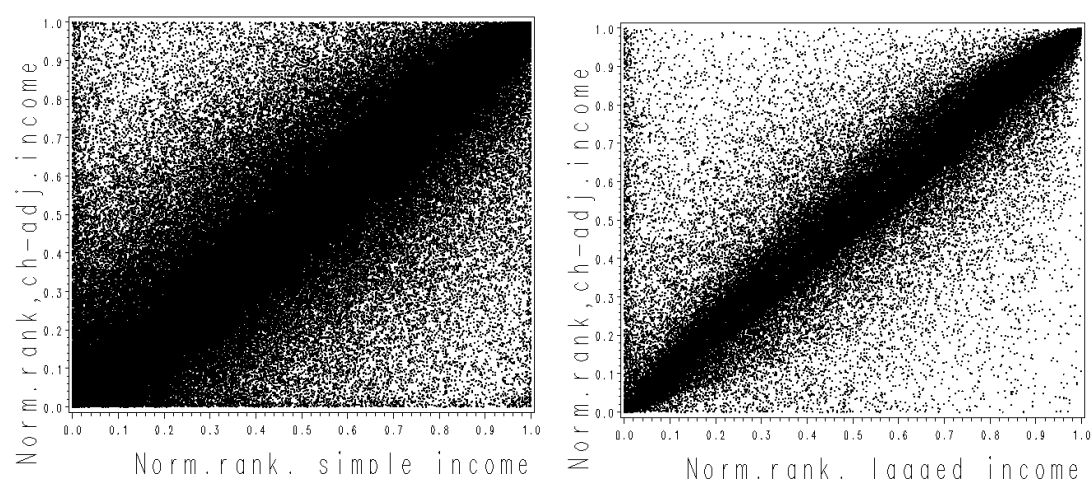
Source: ECHP-UDB 2003

Another way of comparing the three income measures is to determine the relative position of individuals in each of the income distributions and examine to what extent these relative positions differ across the distributions. We define the relative position by the normalised rank of an individuals' income within a certain income distribution (e.g. Böheim and Jenkins, 2000). The normalised rank is obtained by sorting individuals in ascending order of income within each country and for a given year, by determining their rank in this ordering and by dividing this rank by the maximum rank in the distribution.¹⁴

Figure 1 plots the normalised ranks under the change-adjusted income distribution against the normalised ranks under the simple (upper left panel) and the lagged (upper right panel) income distributions respectively.¹⁵ In the resulting scatter plots, the dots situated on the bisector (running from the origin to point [1,1]) represent individuals who maintain the same ranking in both income distributions under consideration. The further away from this bisector, the stronger the re-ranking of the individual's relative income position between the two income distributions. The concentration of individuals around the bisectors in each of the scatter plots indicates that both the simple and the lagged income distribution are positively related to the change-adjusted income distribution. At the same time, however, it is clear that for a considerable number of individuals the normalised ranks deviate substantially in the two plots.

The pattern of these deviations also differs visibly between the two scatter plots. The plot comparing the simple with the change-adjusted ranks shows the largest re-rankings of individuals, both in number and in range. So again, the simple income method yields the largest bias.

Figure 1. Scatter plots of normalised ranks in the change-adjusted versus the simple and lagged income distributions, pooled ECHP (1995-2000).



Source: ECHP-UDB, 2003

One final remark concerns the somewhat denser pattern of dots that can be observed next to the vertical axes running from the origin to the point [0,1] in the two scatter plots in Figure 1. This indicates that a considerable number of individuals situated in the lowest ranks of the simple and lagged income distributions occupy a higher rank in the change-adjusted income distribution.

Table 3. Association between simple, lagged and change-adjusted poverty (percentages), 60% median poverty, 1995-2001, all countries

| | | All sample persons | | | Sample persons with household composition change | | |
|-------------------------------|-------------------|--|------|-------|--|------|-------|
| | | Change-adj. poverty measure | | | Change-adj. poverty measure | | |
| | | non poor | poor | Total | non poor | poor | Total |
| Simple poverty measure | non poor | 78.1 | 5.1 | 83.2 | 75.9 | 5.1 | 81 |
| | poor | 5.4 | 11.4 | 16.8 | 7.9 | 11.1 | 19.1 |
| | Total | 83.5 | 16.5 | 100 | 83.8 | 16.2 | 100 |
| | <i>Kappa [CI]</i> | <i>0.622 [0.620-0.625] (N= 529703)</i> | | | <i>0.554 [0.546-0.562] (N=69263)</i> | | |
| Lagged poverty measure | non poor | 81.3 | 1.1 | 82.3 | 80.1 | 2.6 | 82.7 |
| | poor | 1.1 | 16.6 | 17.7 | 4.8 | 12.6 | 17.3 |
| | Total | 82.4 | 17.6 | 100 | 84.8 | 15.1 | 100 |
| | <i>Kappa [CI]</i> | <i>0.919 [0.918-0.921] (N= 527231)</i> | | | <i>0.713 [0.706-0.720] (N=74246)</i> | | |

Note: [CI] gives the 95% confidence intervals. All figures are longitudinally weighted.

Source: ECHP-UDB 2003

Table 3 presents the association between the poverty measures calculated according to the change-adjusted and the simple and lagged income strategies respectively. The associations are calculated first for all sample persons and then only for sample persons experiencing a household composition change. A first finding is that the marginal poverty distribution estimated among all sample persons is almost identical for the three different poverty measures. About 17 percent of all sample persons are poor according to all measures. Yet, an investigation of the cell frequencies makes clear that not all people classified as poor

according to one measure are also considered poor according to the other one. In line with the other findings in this article, the simple poverty measure resembles the change-adjusted measure least. 5.4 percent is poor according to the simple but not according to the change-adjusted measure; while for 5.1 percent this is the other way around. The dissimilarity index between these two measures, i.e. the total percentage of cases that would need to be reclassified for the two poverty measures to be identical (Agresti, 2002), amounts to 10.5 percent. In contrast, the dissimilarity index for the lagged and change-adjusted poverty measures is substantially lower at 2.2 percent. This finding is confirmed by Cohen's Kappa, which is a measure of inter-rater agreement that equals zero if the level of agreement corresponds to agreement that could be found by chance and that equals one if there is perfect agreement between the measures (Cohen, 1960). The Kappa coefficient of agreement between the simple and the change-adjusted measure is 0.622; indicating substantial agreement; whereas the Kappa for the lagged poverty measure amounts to 0.919, indicating almost perfect agreement (Popping, 1995). The findings for the poverty measures are thus analogous to the results for income measures reported above.

Turning now to the sample of persons experiencing household composition changes during the measurement period¹⁶, we find weaker associations in the two comparisons of poverty distributions, as could be expected. Specifically, the dissimilarity indexes are higher and the Kappas are lower for both association tables in Table 3. In addition, it appears that for this particular sample the misclassifications are no longer symmetrical: more persons are erroneously coded as 'poor' than as 'non poor' according to the simple and lagged poverty measures. Hence, both the simple and the lagged poverty measure overestimate the proportion of poor persons among the individuals with household composition change. This finding confirms the observation from the scatter plots in Figure 2, where a substantial number of individuals was found in the lowest ranks of the simple and lagged income measure, while being higher ranked according to the change-adjusted measure.

5. Inferential analyses with the three poverty measures: modeling poverty entry

5.1. Objective

In this section, we examine how the choice for one of the poverty measures affects the estimation of the effects of certain demographic and labour market related events on the individual's poverty entry risk. The events under study are job loss in the household, child birth and union dissolution. In order to measure the effect of these events accurately, it is important to obtain a match between the timing of the event and the reference period of the poverty measure. The change-adjusted income measure captures the economic well-being of the individual more precisely than the other two measures, because it adjusts for the time lag and for household composition changes. Therefore, the coefficients in the model with change-adjusted poverty will be used as a benchmark.

Which differences in estimated effects can be expected between the poverty measures? For what concerns the simple poverty measure, the problem lies in the correspondence between the timing of the events and the reference period for income. If an event occurred between the start of the calendar year and the interview date d , the effect will be missed by the simple income measure, because the latter is based on previous year's income. This could lead to an underestimation of the true effect of income events. With the lagged poverty measure, a mismatch may arise because the event is measured for the household members present at the interview date d , while the sum of the personal incomes and the equivalence scale are calculated on the basis of the household members present at interview date $d+1$. This can lead to an underestimation of the true effects of demographic events with the lagged poverty measure.

5.2. Model

Discrete-time event history regression analysis is an appropriate technique for this analysis because it models the hazard of an event, in this case the hazard of poverty entry, and is suitable for dealing with events measured in discrete time (annual poverty estimates). The hazard of poverty entry is defined as the probability that an individual will become poor in a certain year, given that this individual was at risk of becoming poor (i.e. not poor) in the previous year (Allison, 2003). Because the dependent variable is a conditional probability, it needs a transformation before it can be included in a regression. We opt for a logit transformation because this is a convenient choice with a straightforward interpretation for many readers (Allison, 1984). Our sample is based on the initial sample persons in the ECHP observed at least on one occasion during the observation window (reference years 1995-2000). It consists of the subsequent person-years in which a sample person is at risk of poverty entry, including the year of the event (person becomes poor) or the year of censoring (participation in panel study ends). In order to avoid dependence among the observations in the dataset, only the person-years until the first poverty entry-event since the start of the observation period are included in the analysis.¹⁷

For the specification of the model, it is important to bear in mind that our primary aim is not to develop the best-specified or most sophisticated model, but rather to assess in a straightforward way the consequences of choosing an inappropriate annual income measure. Therefore, we choose for a rather simple but readily interpretable and comparable model. The model is defined by the following equation (Allison, 2003):

$$(5) \quad \log\left(\frac{P_{it}}{1-P_{it}}\right) = \alpha + \beta_1 jobloss_{d-1,d} + \beta_2 uniondissolution_{d-1,d} + \beta_3 childbirth_{d-1,d} + \beta_4 Htype_d \\ + \beta_5 ageHH_d + \beta_6 ageHH_d^2 + \beta_7 educHH_d + \beta_8 sexHH_d + \beta_9 country_{id} + error$$

In all models, the individual is the unit of analysis and P_{iT} is the conditional probability (or hazard) that individual i becomes poor in year T , given that the individual was not poor in the previous year. The log-odds of the hazard are specified to depend upon three time-varying events of interest, namely job loss, union dissolution and child birth in the household of the individual. All three events took place between the previous interview date $d-1$ and the current date d . Job loss and union dissolution are measured among the persons belonging to the individual's household at the interview date d , whereas child birth is reported for the household at interview date d . Additionally, the effect of the following set of time-varying control variables is estimated: household type ($Htype$), age and age square of the household head ($ageHH$), the highest level of education obtained by the household head ($educHH$), sex of the household head ($sexHH$) and the country of residence of the individual.¹⁸

This model makes a number of simplifying assumptions. First, we assume that the baseline hazard is constant over time. This is a reasonable assumption because the non-poverty spell experienced by most people is left-censored, which makes it impossible to know for how long these people have already been in the non-poor state. In this context, the baseline hazard is not expected to change over time. This is confirmed when estimating a model that does include a year dummy, because this leads to insignificant effects of the year dummy (see Callens, 2005 for similar findings). Second, we do not include any controls for unobserved heterogeneity, because this would make the model less straightforward and because it would impose additional assumptions on the model. Third, for simplicity, we make abstraction of possible problems of endogeneity and simultaneity in the relationship between demographic/labour events and poverty entry (see: Jenkins, 2000, for some further discussion).

5.3. Findings

Table 4 summarizes the findings from the discrete-time event history models for the poverty entry hazard with the simple, lagged and change-adjusted poverty measure. The discussion will focus on the effect of job loss, child birth and union dissolution.

Job loss usually implies an income drop for members of the household. Accordingly, the effect of job loss on the poverty entry hazard is positive and significant for all poverty measures. The effect is largest in the model with the change-adjusted poverty measure. The coefficient for the lagged poverty measure seems to be only slightly underestimated vis-à-vis the change-adjusted measure. The coefficient in the model with simple income is again significantly positive but smaller than in the other models. One reason why we still observe a positive effect even though simple income relates to the previous calendar year T-1, could be that the coefficient partly captures the effect of job losses occurring in the previous calendar year T-1.

The effect of child birth in a household on the hazard of poverty entry is positive and highly significant for all three poverty measures. The coefficient is again largest in the model with the change-adjusted poverty measure. The coefficient in the model with the simple poverty measure is only slightly smaller. This can probably be explained by a similar reasoning as in the case of the effect of job loss. Although the coefficient for the lagged poverty measure is also positive and highly significant, it clearly underestimates the effect of child birth on poverty entry. This can be explained by the fact that the lagged poverty measure applies the equivalence scale at the time of the interview $d+1$ (in calendar year T+1), while the income refers to the calendar year T. One explanation for what is going on is that the higher needs due to the additional child have already been accounted for in the poverty measure for calendar year T-1. Due to this mechanism, some persons will be erroneously categorized as poor in year T-1, and accordingly drop out of the risk set for poverty entry in calendar year T. Consequently, the effect of child birth on the lagged income measure referring to calendar year T is underestimated.

Table 4. Logit coefficients from discrete-time event history models for poverty entry (ECHP, 1995-2000, longitudinally weighted)

| | Simple poverty | Lagged poverty | Change- adjusted poverty |
|---------------------------------|-------------------|-------------------|--------------------------------|
| Intercept | -2.368 *** | -2.311 *** | -2.909 *** |
| Job loss | 0.517 *** | 0.767 *** | 0.840 *** |
| Child birth | 0.403 *** | 0.196 *** | 0.389 *** |
| Partnership dissolution | 0.880 *** | 0.023 | 0.046 |
| Household type | | | |
| Couple without children (Ref) | | | |
| One person household | 0.465 *** | 0.089 ** | 0.188 *** |
| Single parent household | 0.115 *** | 0.087 * | -0.001 |
| Couple with children | 0.106 *** | 0.287 *** | 0.103 *** |
| Other household | 0.048 | 0.260 *** | 0.085 ** |
| Gender household head | | | |
| Male (Ref) | | | |
| Female | 0.264 *** | 0.371 *** | 0.506 *** |
| Education level household head | | | |
| High (Ref) | | | |
| Middle | 0.790 *** | 0.627 *** | 0.913 *** |
| Low | 1.491 *** | 1.337 *** | 1.670 *** |
| Age household head | -0.069 *** | -0.076 *** | -0.066 *** |
| Age ² household head | 0.001 *** | 0.001 *** | 0.001 *** |
| Country | | | |
| United Kingdom (Ref) | | | |
| Austria | -0.199 *** | -0.266 *** | -0.297 *** |
| Belgium | -0.009 | -0.084 | 0.060 |
| Denmark | -0.008 | 0.085 | 0.095 |
| Finland | -0.727 *** | 0.488 *** | -0.727 ** |
| France | -0.232 *** | -0.217 *** | -0.269 *** |
| Greece | 0.325 *** | 0.232 *** | 0.362 *** |
| Ireland | 0.066 | 0.077 | 0.137 ** |
| Italy | -0.014 | -0.171 *** | -0.048 |
| The Netherlands | -0.654 *** | -0.628 *** | -0.588 *** |
| Portugal | -0.217 *** | -0.299 *** | -0.156 *** |
| Spain | 0.199 *** | 0.116 *** | 0.249 *** |
| N | 336386 | 295767 | 283441 |
| Max. rescaled R ² | 0.0723 | 0.069 | 0.0879 |
| -2 Log Likelihood | 155434.6 | 140671.11 | 123288.04 |

Note: *** = $p < 0.001$ ** = $p < 0.01$ * = $p < 0.05$

Source: ECHP-UDB 2003

The effect of union dissolution on the poverty entry hazard diverges substantially between the different poverty measures. In the model with the simple poverty measure, the effect of union dissolution is positive, relatively large and highly significant. In the model with lagged poverty, by contrast, the coefficient estimate is not significantly different from zero. One could argue that this is due to the problems with the measurement of lagged income. However, in the model with the change-adjusted measure, we find a similar non-significant effect of union dissolution on poverty entry. This suggests that the effect of union dissolution found in the model with the simple poverty measure is an overestimation, and probably originates because the simple poverty measure is based on the income of the previous calendar year T-1. This might lead to an underestimation of the income for the partner who earned a low –or no– personal income before the union dissolution. The reason is that income adjustments following union dissolution are neglected in the calculation of the simple income measure.¹⁹

A last observation concerns the fit of the different models. Overall, the change-adjusted model leads to the largest explained variance and the lowest -2 LogLikelihood value. The explained variance is rather low in all three models under study, indicating that better models could probably be fit, but the search for the best model is beyond the scope of this article.

6. Conclusion

Annual income in surveys can only be reported after the reference year has elapsed, so by definition it is provided with a time lag. Ignoring this time lag may lead to inaccurate estimations of the household-based income for individuals. In the literature, several methods can be distinguished for dealing with this time lag. The article at hand has reviewed and evaluated the most frequently encountered methods in the context of household panel data research. The first

method has been called the simple income method. We have argued that it does not provide an appropriate solution because it takes the income relating to the previous year as a proxy for current year's income, so basically it ignores the time lag. The second method that has been evaluated has been called the lagged income method because it tries to circumvent the time lag problem by using the income reported in the following year as the current year's income. In doing so, however, it relies on the resources and needs structure of the household in which the individual resides after the calendar year has elapsed. In case of changes in the household composition, this leads to bias in the lagged measure as well.

With a view to these problems, we have proposed an alternative income method in this article, namely the change-adjusted income method. It uses retrospective information about the timing of household composition changes to identify the households to which an individual belonged during the reference year. On the basis of this information, it calculates the household income for each of these households; divides it by the correct equivalence scale and takes a weighted sum to obtain an annual amount. In this way, the change-adjusted measure does not only correct for the time lag, but also for household composition changes during the relevant period.

In the empirical part of the article, we have compared the three income methods in different cross-sectional and longitudinal analyses. The main conclusion is that the simple and lagged methods tend to induce bias in income and poverty analyses. However, both measures do not perform equally badly, and the bias in the findings also depends on the specific analysis researchers want to perform.

The simple income method clearly performs worst and displays the largest bias in its estimated medians, its income distribution, its relative ranks and its income poverty estimates. The lagged income method performs better in this respect, but there are still significant differences with the change-adjusted measure. In models predicting the hazard of entering poverty, the model with lagged poverty shows a slight underestimation of the effect of child birth,

whereas the model with simple poverty slightly underestimates the effect of job loss and clearly overestimates the effect of union dissolution.

Generally, the findings in this article show that it is advisable to adjust for the time lag inherent to annual income survey data and for household compositions during the year. However, the available correction strategies come at varying costs, and these should be taken into account. Firstly, when the data are provided in a panel survey, researchers should consider whether it is feasible to lose one or more waves for the calculation of income. Secondly, attention should be paid to nonresponse patterns. A measure that uses the income information from the next wave will be more affected by dropout from the panel. Finally, the accuracy of the change-adjusted income method will depend on the quality of the available retrospective information on household composition changes. It is up to the researcher to use the evidence in this article as a guide for weighing up the costs and benefits associated with the different time lag correction strategies.

References

- Agresti, Alan, Categorical Data Analysis, New Jersey, 2002.
- Allison, Paul. D., Event History Analysis. Regression for Longitudinal Event Data, Sage, London, 1984.
- Allison, Paul. D., Survival analysis using the SAS system: a practical guide, SAS Institute, Cary, 2003.
- Apospori, Eleni and Jane Millar, Eds. (2003). The dynamics of social exclusion in Europe : comparing Austria, Germany, Greece, Portugal and the UK, Elgar.
- Atkinson, Tony, Bea Cantillon, Eric Marlier and Brian Nolan, Social indicators : the EU and social inclusion, Oxford university press, Oxford, 2002.
- Bane, Marie J. and David Ellwood, "Slipping into and out of poverty: The dynamics of spells." Journal of Human Resources, Series 21, 1-23, 1986.
- Barnes, Matt, Christopher Heady and Sue Middleton, Eds. (2002). Poverty and social exclusion in Europe, Elgar.
- Böheim, René and Stephen Jenkins, "Do current income and annual income measures provide different pictures of Britain's income distribution?" DIW Discussion Paper, 29, 2000.
- Breen, Richard and Pasi Moisio, "Poverty dynamics corrected for measurement error." Journal of Economic Inequality, Series 2, 171-191, 2004.
- Cabrero, Gregorio R., "Poverty in Spain: a critical analysis based on the European Community Household Panel (ECHP)." Journal of European Social Policy, Series 13, 93-94, 2003.
- Callens, Mark, "Poverty dynamics in Europe. A multilevel discrete-time recurrent hazard analysis." EPUNet-2005 Conference: the conference of the European Panel Users Network, Colchester, 2005.

- Cantillon, Bea, Rudi Van Dam, Karel Van den Bosch and Bart Van Hoorebeeck, "The impact of the reference period on measures of household income from surveys." Berichten, 27, 2003.
- Cohen, Jacob, "A coefficient of agreement for nominal scales." Educational and Psychological Measurement, Series 20, 37-46, 1960.
- Dennis, Ian and Anne-Catherine Guio, "Poverty and social exclusion in the EU after Laeken - part 1", Statistics in focus, 8/2003, 8, 2003.
- Dennis, Ian and Anne-Catherine Guio, "Poverty and social exclusion in the EU", Statistics in focus, 16/2004, 12, 2004.
- European Central Bank, "Statistical data warehouse. HICP, other prices and costs." 2007, Retrieved 5 April 2007 from <http://sdw.ecb.int/reports.do?currentNodeId=100000181>.
- Eurostat, "Imputation of income in the ECHP", DOC. PAN 164/2002-12, 2002.
- Eurostat, "ECHP UDB. Description of variables. Data Dictionary, Codebook and Differences between Countries and Waves", Doc PAN 168/2003-06, 2003.
- Fouarge, Didier and Richard Layte, "Welfare Regimes and Poverty Dynamics: The Duration and Recurrence of Poverty Spells in Europe." Journal of Social Policy, Series 34, 407-426, 2005.
- Jenkins, Stephen P., "Modelling household income dynamics." Journal of Population Economics, Series 13, 529-567, 2000.
- Kuchler, Birgit and Jan Goebel, "Incidence and intensity of smoothed income poverty in European countries." Journal of European Social Policy, Series 13, 357-369, 2003.
- Layte, Richard and Christopher T. Whelan, "Moving in and out of poverty - The impact of welfare regimes on poverty dynamics in the EU." European Societies, Series 5, 167-191, 2003.
- Mejer, Lene and Clemens Siermann, "Income poverty in the European Union: Children, gender and poverty gaps", Statistics in focus, 12/2000, 8, 2000.

- Popping, R., "User's Manual AGREE 6.0, Computing Agreement on Nominal Data, ProGAMMA", 1995.
- Schlotzhauer, Sandra D. and Ramon C. Littel, SAS® System for Elementary Statistical Analysis, The SAS Institute, Tervuren, 1997.
- Smeeding, Timothy M. and Daniel H. Weinberg, "Toward a uniform household income definition." Review of Income and Wealth, Series 47, 24, 2001.
- Walker, Robert, Poverty Dynamics: Issues and Examples, Avebury, Aldershot, 1994.
- Whelan, Christopher T., R. Layte and B. Maitre, "Persistent income poverty and deprivation in the European Union: An analysis of the first three waves of the European Community Household Panel." Journal of Social Policy, Series 32, 1-18, 2003.
- Whelan, Christopher T., Richard Layte and Bernard Maitre, "Multiple deprivation and persistent poverty in the European Union." Journal of European Social Policy, Series 12, 91-105, 2002.
- Whelan, Christopher T., Richard Layte, Bernard Maitre and Brian Nolan, "Income, deprivation, and economic strain. An analysis of the European Community Household Panel." European Sociological Review, Series 17, 357-372, 2001.

¹ Households are difficult to follow over time because they can change or split.

² This is based on the reasonable assumption that resources and needs are shared within households.

³ See Jenkins (1991) for a discussion of this assumption.

⁴ Equivalence scales are commonly applied to adjust for differences in household needs due to differences in household composition and size while taking into account economies of scale.

⁵ At least if these time-varying covariates are not measured retrospectively.

⁶ According to the simple income method, household information is more in line with the reference period of income because both the equivalence scale and income measured at interview date d are assumed to refer to the current calendar year T .

⁷ With a view to this difficulty with the lagged income strategy, some authors have proposed to lag the household income information measured at time $d+1$ ($Y_{h_{d+1},T}$) by one year but to match it to the equivalence scale measured at time d (E_{h_d}) instead of the one measured at time $d+1$

($E_{h_{d+1}}$) (Breen and Moisisio, 2004; Layte and Wheland, 2003): $Y_{i_h,T}^* = \frac{Y_{h_{d+1},T}}{E_{h_d}}$. However, the bias in

this matched income is possibly even more severe than in the lagged income method because the equivalence scale E_{h_d} is based on a different household composition than the sum of the personal incomes $Y_{h_{d+1}}$. Empirical analyses for the matched income method have not been included in this article, but are available from the authors upon request.

⁸ Special thanks go to two anonymous reviewers who made us aware of the possibilities of this measure. Note that one could come up with other formulas that stick to the household level.

However, as we wanted to focus on the individual as the unit of analysis (see footnote 1), this

measure is the most correct reflection for the household-based income of the individual, and can thus be used as a benchmark to assess the bias arising from the other two measures.

⁹ In Austria and Luxembourg the ECHP panel study started only in 1995 and in Finland in 1996. Also for these countries, the first participation year is omitted from the analyses, i.e. 1995 is omitted for Austria and Luxembourg and 1996 for Finland.

¹⁰ All annual net household incomes presented in this article are composed of the net personal incomes of the household members aged 16 and over, as well as of a number of household income components. The personal incomes consist of: total net income from work, total net private income and the total net amount of social insurance benefits. The household income components are: rental income/assigned property, housing allowance, and social assistance. The modified OECD scale attributes a weight of 1 to the first adult in the household, and adds a weight of 0.5 for each additional adult and a weight of 0.3 for each additional child.

¹¹ Note that this operationalisation is based on the assumption that individuals spread their annual personal income equally over the twelve months of a calendar year. This assumption could have been avoided if personal incomes for each month t during calendar year T would have been available in the data, but unfortunately this was not the case.

¹² Since the lagged and change-adjusted income measures rely on information measured in the following wave, only those observations are considered with an income value for the two subsequent waves 1995-1996, 1997-1998 or 2000-2001 respectively. Additionally, all observations with item nonresponse on one of the income measures have been omitted. To correct for attrition occurring between the two neighbouring waves, the ECHP base weight of respectively 1996, 1998 and 2001 is applied.

¹³ This test was chosen because the normality assumption was not met according to the Kolmogorov-Smirnov test (Schlotzhauer and Littel, 1997).

¹⁴ For reasons of comparability, only individuals without missing values on each pair of income measures being compared are included in the calculation of the normalised ranks.

¹⁵ As we found a consistent pattern across countries and years, we display pooled results for all countries and for all years from 1995 until 2000 in Figure 1.

¹⁶ For the simple poverty measure, household composition changes between $d-1$ and the end of calendar year T are taken into account. For the lagged poverty measure, household composition changes in the period between the beginning of calendar year T and $d+1$ are taken into account.

¹⁷ Some techniques can deal with repeated events, for example fixed or random effects specifications, but these will not be treated here in order not to complicate the model (in view of its purposes) and in order not to introduce additional assumptions.

¹⁸ All control variables are measured for the household at interview date d .

¹⁹ These income adjustments could result from e.g. alimony, social benefits, a new job or changes in the working pattern.