METHODS AND NOMENCLATURES

ISSN 1725-0056
Cat. No. KS-BF-03-001-EN-N

2003 EDITION

The EC Household Panel "Newsletter" (01/02)

3

EUROPEAN COMMISSION

THEME 3
Population and social conditions

Europe Direct is a service to help you find answers to your questions about the European Union

New freephone number:

00 800 6 7 8 9 10 11

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Luxembourg: Office for Official Publications of the European Communities, 2003

ISSN 1725-0056

The EC Household Panel "Newsletter" (01/02)

Petra Lehmann - Christine Wirtz

At the Lisbon Summit in March 2000, the European Council invited 'the Commission to draw up an annual synthesis report on progress in areas related to employment, innovation, economic reform and social cohesion on the basis of Structural Indicators'. Since the conclusions of the Nice European Council in December 2001, social statistics, particularly in the area of 'at risk of poverty and social exclusion', have been even higher on the political agenda. This means that statistics in this important area are under a lot of scrutiny, at Member State level, as well as at European level.

Four of the structural indicators published in the synthesis reports since 2000 are derived from the ECHP. They are:

- distribution of income (S80/S20 ratio);
- at-risk-of-poverty rate (before and after social transfers);
- persistent at-risk-of-poverty rate; and the
- gender pay gap

Various other indicators related to social statistics that also use the ECHP as a data source are constantly being used for research and policy purposes by researchers and politicians inside and outside the Commission.

This high-level publication and usage of ECHP results in this sensitive policy area, together with the first ECHP based report on poverty and social exclusion in the EU attracted much attention regarding the quality of the ECHP data. Several countries became therefore increasingly aware of certain deficiencies in their data. A number of countries even revised their micro-data, some of them back to the very start of the ECHP in 1994. The common methods regarding the imputation of income data, as well as the weighting procedures also showed some weaknesses, which had to be addressed.

The aim of this newsletter is therefore to present three important issues regarding the ECHP that have an impact on the quality of the data, i.e. the development of sample sizes in the ECHP since the beginning of the survey, the subject of 'imputation of income', as well as information on methodological changes applied to the waves 1-6 issue of the ECHP UDB. In the section on 'sample sizes', the newsletter aims to give an overview over the achieved sample sizes and the attrition rate in the ECHP from 1994 to 1999. The second section covers 'imputation', which is an important aspect of a panel survey and this newsletter will describe the current situation in the ECHP regarding imputation of income on the basis of a number of income components. The final section will explain the modifications in the weighting procedure used in the ECHP applied to the waves 1-6 issue of the ECHP, as well as some explanations on how the ECHP deals with 'within household non-response'.

1. SAMPLE SIZE

The design of the ECHP as a household panel survey means that the same set of individuals and households is followed over time, which makes it subject to sample attrition, i.e. the loss of respondents over the life of the survey. Non-response occurs at two distinct stages: at the household interview stage, and at the stage of personal interviews. In the ECHP, the tracing rules clearly define the sample of persons to be followed-up from one wave to the next. Whatever happens to these persons, some information should be recorded in the ECHP register files. If a person is for example 'out-of-scope' (died, institutionalised, moved out of the EU) this should be recorded. For persons still 'in-scope', their household should be interviewed. Reasons for why an in-scope person leaves the sample (refusal, not found) should also be recorded.

Tables 1 and 2 present some information on the evolution of sample sizes for 'interviewed households' and 'interviewed persons' (only persons aged 16 and over are eligible for a personal interview in the ECHP). The ECHP started in 1994 in the then 12 Member States. Austria joined the project in 1995 and Finland in 1996. Data for Sweden is available from 1997 onwards. The data is derived from the Swedish Living Conditions Survey[1] and the variables are transformed into ECHP format. In the UK, Luxembourg and Germany, the ECHP survey was stopped in 1997. Data from an existing national panel was transformed into the ECHP format corresponding to the ECHP variables to provide data for all waves.

Table 1: Number of interviewed households

	B	DK	D ECHP	D SOEP	EL	E	F	IRL	I	L ECHP	L PSELL	NL	A	P	FIN	S	UK ECHP	UK BHPS	EU*
Number of interviewed households																			
1994	3490	3482	4968	6207	5523	7206	7344	4048	7115	1011	:	5187	:	4881	:	:	5779	5124	61273
1995	3366	3223	4688	6336	5220	6522	6722	3584	7128	962	2978	5110	3380	4916	:	:	4548	5025	61017
1996	3210	2955	4593	6259	4907	6267	6600	3173	7132	933	2472	5179	3292	4849	4139	:	3775	4994	62670
1997	3039	2745	:	6163	4604	5794	6176	2945	6713	:	2654	5049	3142	4802	4106	5385	:	4943	68260
1998	2876	2512	:	5962	4211	5485	5866	2729	6571	:	2523	4963	2960	4716	3920	5308	:	4966	65568
1999	2712	2387	:	5847	3986	5418	5610	2378	6370	:	2552	5023	2815	4683	3822	5250	:	4911	63764
… change as a percentage of previous wave																			
95/94	-3.6	-7.4	-5.6	2.1	-5.5	-9.5	-8.5	-11.5	0.2	-4.8	:	-1.5	:	0.7	:	:	-21.3	-1.9	-0.4
96/95	-4.6	-8.3	-2.0	-1.2	-6.0	-3.9	-1.8	-11.5	0.1	-3.0	-17.0	1.4	-2.6	-1.4	:	:	-17.0	-0.6	2.7
97/96	-5.3	-7.1	:	-1.5	-6.2	-7.5	-6.4	-7.2	-5.9	:	7.4	-2.5	-4.6	-1.0	-0.8	:	:	-1.0	8.9
98/97	-5.4	-8.5	:	-3.3	-8.5	-5.3	-5.0	-7.3	-2.1	:	-4.9	-1.7	-5.8	-1.8	-4.5	-1.4	:	0.5	-3.9
99/98	-5.7	-5.0	:	-1.9	-5.3	-1.2	-4.4	-12.9	-3.1	:	1.1	1.2	-4.9	-0.7	-2.5	-1.1	:	-1.1	-2.8
… percentage change wave 6/ wave 2																			
99/95	-19.4	-25.9	:	-7.7	-23.6	-16.9	-16.5	-33.6	-10.6	:	-14.3	-1.7	-16.7	-4.7	:	:	:	-2.3	4.5

Source: ECHP UDB version of December 2002

** The EU figure here includes the sample size of each available country per year (for the UK and Luxembourg it includes the ECHP data for 1994-1996; for Germany it includes the SOEP data for all years[2]). Therefore, as more countries joined the ECHP after 1994, the sample size for the EU as a total increased.*

[1] *The 'Swedish Living Conditions Survey' is a cross-sectional survey, not a panel survey*

[2] *For the calculation of income indicators based on the ECHP at Eurostat, the same procedure applies, i.e. for the UK and Luxembourg for 1994-1996 ECHP data are used, for Germany SOEP data are used for all the years*

Table 1 compares the number of households interviewed in each wave. Each wave, the sample falls short of that of the previous year due to non-contact, non-response, failure to follow-up the sample cases for other reasons, households ceasing to exist, etc. These losses are compensated to some extend by the inclusion of new (split-off) households coming into the sample as a result of the movement of sample persons. The figures summarise the net effect of these factors.

Overall, the wave 6 sample for the 13 countries for which data is available since 1995 is 86.1 % as large as the wave 2 sample (63510 households in 1995 as compared to 54692 in 1999). For this calculation, the sample sizes of only those countries were used for which data was available in 1995 and after. It therefore uses the sample sizes of the national panels for Germany, Luxembourg and the UK.

Table 2: Number of interviewed persons

	B	DK	D ECHP	D SOEP	EL	E	F	IRL	I	L ECHP	L PSELL	NL	A	P	FIN	S	UK ECHP	UK BHPS	EU*
										Number of interviewed persons									
1994	6710	5903	9490	12233	12492	17893	14333	9904	17729	2046	:	9407	:	11621	:	:	10517	8915	130788
1995	6454	5503	9002	12542	12271	16263	13306	8531	17780	1968	6785	9151	7437	11858	:	:	8386	8681	131450
1996	6145	4994	8746	12295	11602	15640	13051	7487	17736	1915	5611	9277	7271	11702	8173	:	6940	8877	134228
1997	5741	4628	:	12059	10968	14819	12143	6868	16594	:	5802	9089	6999	11625	8068	9597	:	8865	143865
1998	5339	4187	:	11562	9985	13779	11209	6324	15934	:	5410	8826	6561	11412	7381	9461	:	8764	136134
1999	5021	3983	:	11288	9574	13104	10682	5451	15401	:	5291	8917	6246	11250	7110	9314	:	8601	131233
										… change as a percentage of previous wave									
95/94	-3.8	-6.8	-5.1	2.5	-1.8	-9.1	-7.2	-13.9	0.3	-3.8	:	-2.7	:	2.0	:	:	-20.3	-2.6	0.5
96/95	-4.8	-9.2	-2.8	-2.0	-5.5	-3.8	-1.9	-12.2	-0.2	-2.7	-17.3	1.4	-2.2	-1.3	:	:	-17.2	2.3	2.1
97/96	-6.6	-7.3	:	-1.9	-5.5	-5.2	-7.0	-8.3	-6.4	:	3.4	-2.0	-3.7	-0.7	-1.3	:	:	-0.1	7.2
98/97	-7.0	-9.5	:	-4.1	-9.0	-7.0	-7.7	-7.9	-4.0	:	-6.8	-2.9	-6.3	-1.8	-8.5	-1.4	:	-1.1	-5.4
99/98	-6.0	-4.9	:	-2.4	-4.1	-4.9	-4.7	-13.8	-3.3	:	-2.2	1.0	-4.8	-1.4	-3.7	-1.6	:	-1.9	-3.6
										… percentage change wave 6/ wave 2									
99/95	-22.2	-27.6	:	-10.0	-22.0	-19.4	-19.7	-36.1	-13.4	:	-22.0	-2.6	-16.0	-5.1	:	:	:	-0.9	-0.2

Source: ECHP UDB version of December 2002

** The EU figure here includes the sample size of each available country per year (for the UK and Luxembourg it includes the ECHP data for 1994-1996; for Germany it includes the SOEP data for all years[2]). Therefore, as more countries joined the ECHP after 1994, the sample size for the EU as a total increased.*

Overall, the wave 6 sample for the 13 countries for which data is available since 1995 is 83 % as large as the wave 2 sample (124704 persons in 1995 as compared to 103559 in 1999).

Both for 'interviewed households', as well as for 'interviewed persons' it can be seen that attrition tends to be lowest in those countries, where a panel survey has already been in existence for a number of years, i.e. the German and UK national panels (SOEP and BHPS respectively), as well as the Dutch ECHP, where the data also comes from a long running national panel. One exception is Portugal, where attrition was low throughout, despite having started the ECHP as a new survey in 1994.

2. IMPUTATION

The following section outlines the general approach followed in the ECHP regarding the imputation of item non-response in relation to income. Total household income in the ECHP is made up of a large number of income components, like for example capital income, unemployment related benefits, etc. For some components detailed sub-components are collected (e.g. unemployment insurance benefits, unemployment assistance, etc.). It is not acceptable to reject a case if the information is incomplete, as that would result in the loss of much valuable information. Hence, it is essential to impute missing values in the income variables where that can reasonably be done. Another important characteristic is that the income variables form a set of variables in which there is an interdependence between all the components. Therefore it is not meaningful to impute individual components separately. The imputation approach used for the ECHP consists of a multivariate model involving a multiple regression sequence.

The variables are divided into two types of variables, the auxiliary variables, which are used to impute the others, and the target variables, which are the subject of the imputation. At the beginning of the imputation process, the auxiliary variables are those relating to demographic characteristics (e.g. sex, age, etc. at *person* level; region, household size, household type, etc. at *household* level), and to the labour force characteristics. The target variables are essentially elementary components of income (capital income, income from self-employment, etc.) or aggregated components of income (total income from work or unemployment related benefits, etc.).

For the imputation process, the target variables are arranged in a sequence, starting with those with the smallest proportion of (or with no) missing values. Each target variable is imputed sequentially, using all the variables above it in the sequence order for which all information is available (or has been previously imputed). Once a variable with missing values has been imputed, it is moved from the second set to the first, i.e. it will be used as an auxiliary variable for imputation of the next variable on the 'list'.

After all variables in the list have been dealt with , the process is started again with the first variable in the target set, but this time using all the other variables as predictors, using for each the given or the most recently imputed value. The process is performed for each variable in turn, and is repeated sequentially.

Households for which the income information available is so incomplete that no reasonable basis exists for imputation are excluded, however this concerns only a small proportion of households in the ECHP.

Imputation with the IVE software

The imputation of amounts is carried out using the impute module of the IVE software ('Imputation and Variance Estimation Software', by the Survey Research Centre, Institute for Social Research of the University of Michigan). The software and documentation of this software is available under 'http://www.isr.umich.edu/src/smp/ive'

Actual imputation of income

In the ECHP, information on income from a particular component is mainly collected using the following sequence of questions:

(i) whether or not any income has been received from that source (filter question);

(ii) where applicable, the number of months received;

(ii) where applicable, the amount received per month.

The imputation procedure used in the ECHP for income components collected with this sequence of questions is the following:

(i) whether or not any income has been received from a specified source

When this answer is missing, this item is set to 'yes' or 'no' depending on component specific rules. A filter question may be set to 'yes', based on other information collected during the interview. If, for example, a person has been unemployed for at least one month in the previous year, and the information on whether unemployment benefits have been received is missing, the first four components related to unemployment benefits[3] would be set to 'yes', while the last component (other unemployment related benefits) would be set to 'no'. Another example would be the case of a household with children below 16 years of age. In this case, a missing answer to the question whether family related allowances are obtained would be replaced by 'yes', otherwise it would be set to 'no'.

(ii) where applicable, the number of months is missing

If the number of months are missing, they are imputed in two steps. Firstly the 'calendar of activities' is used. For example if a person had been unemployed for 3 months and the number of months corresponding to unemployment related benefits is missing, then it would be set to 3. For those persons with missing calendar information, or for variables that are not related to activities (e.g. deserted wives allowance), the number of months is imputed by applying a random number of months, which is based on the distribution for responses (including imputations derived from calendar information).

(iii) where applicable, the amount received per month

A. If for a unit (household or person) the amount is unknown, the corresponding amount from the most recent previous wave, whether imputed or declared, is forwarded.

- For the initial wave of a country information from the following wave is copied, however, this is never imputed but only declared.
- Even if a unit was interviewed before, it can happen that the specific income component was not obtained. Thus the amount is still missing after this operation.

B. Some components are imputed as such (i.e. with the same detail as they are collected), others are aggregated (i.e. some components are grouped together) and the aggregate is imputed. For missing components, lower and upper limits are specified and the imputed value will fall inside this predefined interval.

A detailed description of the imputation procedures used in the ECHP is available as DocPan 164 from the ECHP Circa site: http://forum.europa.eu.int/Public/irc/dsis/echpanel/home

[3] *These are: unemployment insurance benefit, unemployment assistance, training/retraining allowance, and placement, resettlement, and rehabilitation benefits*

3. MODIFICATIONS APPLIED TO THE WAVES 1-6 ISSUE OF THE ECHP UDB

This section focuses on the changes undertaken in the methodological procedures used in the ECHP and the reasons for those changes.

Until now, whenever a new version of the ECHP UDB was issued, important changes referring to all years were made. However, the changes were mostly limited to some national subsets. This time, the modifications related to certain procedures, e.g. the weighting procedure, were applied to all countries. In general, two substantial modifications in the production of the ECHP UDB led to important changes in the main indicators derived from these data. The main changes as compared to earlier versions are:

- an improved weighting procedure of ECHP data, in order to avoid extreme weights; and
- the application of a new method to adjust for 'within household non-response'.

The ECHP samples and the weighting procedures

The ECHP started off in all the participating EU Member States with a sample of households. All people living in these households were considered to be 'sample persons'[4], i.e. a member of the initial sample, that would be followed up during the life of the ECHP.

For the initial wave of each national ECHP subset, the weights which would be used for analysis were derived. The calculation of these weights took account of sample selection probabilities, non-response and external information, such as the distributions known from large-scale surveys or registers.

Due to the fact that, although a sample of households was selected, results referring to the population of individuals in the Member State are also needed, weights were calculated for both units of analysis, persons and households.

In subsequent waves, all households consisting of at least one 'sample person' were to be followed up, i.e. all its members should have been interviewed. For each subsequent wave, new weights were calculated taking into account sample attrition. Weights from the previous wave were used as starting weights. The 'new weights' were adjusted for non-response and calibrated to take into account external information.

Up to wave 5, to calculate the 'current' weight of a sample person in the ECHP, the persons weight from the previous wave was used as a start and adjusted by the non-response probability of persons with similar characteristics. Subsequently the average weight of all sample persons in a household was calculated, and this weight was calibrated according to external control distributions. The aim of the final weight is for the overall distribution of sample person to be as similar as possible to the distribution in the actual population.

Due to the fact that for cross-sectional analysis not only sample persons, but also non-sample persons living in the same households should be analysed, cross-sectional weights for all persons were calculated by sharing the weights of the sample persons equally among all household members. Each household also received a cross-sectional household weight, which was proportional to the cross-sectional weights of the household members.

[4] *A child born to a sample mother is also considered to be a sample person*

Due to the long duration of the ECHP, and without supplementing the sample to make up for selective attrition, the weights in the latter waves increasingly showed extreme values. Individuals living in households with very specific characteristics, especially in large households, usually obtained very high weights. It has turned out that extreme weights have mainly been introduced through calibration. For previous issues of ECHP data, strong emphasis was given to the exact calibration according to several external distributions. Some of these external characteristics could not be derived in detail from the ECHP data, e.g. 'number of economically active people in a household', which was difficult to obtain when the activity status was not known for all household members. Other characteristics could be easier derived from the ECHP, e.g. household size. However, the actual definition of 'household'[5] sometimes differed between the ECHP and external sources.

In its meeting of 27 May 2002, the ECHP Working Party decided to review the above described weighting procedure. This has been done for the waves 1-6 issue. This review concerns the updating of all weights from the second wave onwards. The weighting has been changed by eliminating certain external distributions from the calibration. In addition, the changes of the weights from one wave to the next have been limited by trimming the adjustment factors. These changes now result in less extreme weights.

Adjustment for within household non-response

Another aspect which has been changed is the 'adjustment for within household non-response'. For analysis on income and poverty, it is important to know the total income of a household. As most income components are collected from the individuals living in the household, a missing questionnaire from one individual yields incomplete household income.

The importance of within household non-response varies remarkably between countries: in 1997, from 21 households (0.5%) with one missing personal interview in Greece, up to 522 households (10.3%) in the Netherlands or 272 (8.9%) in Belgium.

Up to the waves 1-5 issue of the UDB, individual non-response has been adjusted for by multiplying, at household level, all income components with the 'within household non-response inflation factor'. This inflation factor was constructed in order to provide a reasonable estimate of total household income. However, this calculation disturbs the information at component level. It also underestimates the household income when the main income earner is the non-respondent.

Eurostat therefore developed a new way of adjusting for within household non-response in close co-operation with the national data collection units (NDUs). The new method estimates the amount of income which is missing, as opposed to the share of income which had been estimated by the old method.

Some countries, namely the UK, Ireland and Finland, decided to provide the estimates for within household non-response themselves. In this case, the values provided are simply added to the income collected during the interviews.

[5] A household is sometimes defined according to the dwelling concept, i.e. all people living in the same dwelling, while the ECHP defines a household as all people having common living arrangements. The household according to the 'common living arrangements' concept is a sub-unit of the 'common dwelling' concept

For the other national subsets of the ECHP UDB, a common method for estimating the amount has been applied. The missing income is estimated using other information from interviews (personal income from previous year and/or total household income from current/previous year). This is done as follows:

1. If the person was interviewed in the previous wave: the income from the previous wave is copied and added to the household income.
2. For households with no change in composition (nobody moving out or in, died or new born), and with a completed household interview in the previous wave, the difference D between the annualised current net household income from the previous wave and the sum of all the income collected in the current wave (including the income copied from the previous wave for non-responding persons) is calculated. If the value is positive, this amount is added to the household income.
3. For households with a change in composition (at least one person moving out or in, died or new born), or without a completed household interview in the previous wave, the similar procedure as described under (2) is applied, using the current net income from the current waves instead of the previous wave.

➢ ACCESS TO ECHP DATA

ECHP: at a glance

The European Community Household panel (ECHP) is an EU-specific survey that provides comparable micro-data concerning the living conditions of private households and individual persons throughout the EU Member States. Covering income, employment status, housing, healthcare, education, poverty and social exclusion, and other social issues, these EU-wide social data are collected in a harmonious structure that enables them to be efficiently integrated into comparison studies. Created and maintained by the Statistical Office of the European Communities (Eurostat), in close consultation with the Member States, the ECHP database offers researchers a unique resource for comparing data at the micro-level throughout the EU.

ECHP UDB: your 'keyhole' on EU social statistics

Direct access to the original data is restricted. Due to the strong demand for access by researchers and other users, however, Eurostat has developed an 'anonymised' version of the database that complies with data-confidentiality requirements: the European Community household panel users' database or ECHP UDB. The ECHP UDB is a user-friendly database and accessible to those who sign a research contract with Eurostat, which stipulated strict conditions of data access and use.

Scope of the ECHP UDB

The ECHP UDB micro-data are structured in a series of eight survey 'waves', from Wave 1 (1994) to Wave 8 (2001). Countries covered in Wave 1 were the 12 EU Member States at that time, with subsequent waves including data for new countries joining the EU (Austria – Wave 2: Finland – Wave 3). From 1997, comparable data extracted from the Swedish living conditions survey are also available.

Restricted access

Only those universities and other organisations undertaking research work which enter into a contractual arrangement with Eurostat may obtain access to ECHP micro data. There is no possibility for individual persons to acquire the ECHP UDB CD-ROM directly. Requests to sign an **ECHP research contract** must come from the organisation with which the interested person is studying or working. All contractors must strictly adhere to the terms and usage conditions set out in the contract. Sample conditions are as follows.

The contractor must ensure that the files contained in the ECHP UDB are used exclusively for the research purposes clearly specified in the contract, excluding, in particular, any possible administrative use.

The contractor will be required to take all the necessary regulatory, administrative, technical and organisational measures to ensure that none of these data are distributed to third parties, and that there will be no attempt to identify, by any means whatsoever, any individual statistical unit. In particular, the ECHP UDB shall not be connected with other data sets from any other source.

A copy of all the data, i.e. the whole CD-ROM containing the ECHP UDB, can be made only if both the contract allows for this and explicit authorisation of the contractor's technical manager is given.

Contracts now cover all the waves of the ECHP, i.e. those currently available as well as those to come later. Those organisations signing an ECHP research contract will initially receive the available data set (e.g. for contracts signed from January 2003 onwards, the 1994-1999 data set).

Updated data sets (1994-2000 and 1994-2001) will be provided as they become available.

Need a closer look?

A description of the ECHP UDB (containing a data dictionary and the codes and labels of the individual variables, as well as the 'anonymisation criteria' that have been applied) is given in the document *ECHP UDB – A description of variables,* available from the following Internet address:

http://forum.europa.eu.int/Public/irc/dsis/echpanell/home

Document *PAN170: Research contracts using ECHP data*, also available from this address, provides a lost of research projects/publications based on the ECHP data. It is constantly updated.