EUROPEAN COMMISSION
EUROSTAT

Directorate E: Social statistics
Unit E-2: Living conditions

DOC. PAN 165/2003-06

CONSTRUCTION OF WEIGHTS IN THE ECHP[1]

[1] The weighting procedure described in this document has been applied at Eurostat for all national subsets of the ECHP UDB, except for the Swedish, the Luxembourgish PSELL and the British BSPH data. For these three datasets, weights have been provided by the National Data Collection Unit.

June 2003

Table of Contents

1. INTRODUCTION

In the ECHP UDB files, weights are included for households and persons. These weights are calculated taking into account the sample design - this is reflected by the design weights - and characteristics of persons and households. The weights are calibrated to reflect the structure of the population.

This document describes the weighting procedures that have been implemented for calculating weights of in the ECHP. It does not describe explicitly how these weights are to be used.

For a short description of sample weights and on how to use these weights, please see the ECHP UDB manual - Doc.PAN.168, annex 2.

2. THE ALGORITHM

In general, starting weights have to be available. These are modified from one wave to the next in inverse proportion to response probabilities and similar measures.

Generally, response probabilities and similar measures are defined from binary (yes-no) indicators using logit or probit regression controlling for basic household and personal characteristics.

2.1. Household and personal characteristics

In a first step, variables describing households and persons are copied and derived from different data sets. All the variables necessary for the calculations are described in this section.

2.1.1. Weighting status of a person in the current wave (variable name: PWSTAT)

This information is available from the link-file of the ECHP UDB. The variable PWSTAT) has values 1 (resident, i.e. the person is member of an interviewed household in this wave), 2 (person is not a member of an interviewed household in this wave, but it is still in scope of the survey), 3 (person is not a member of an interviewed household in this wave, and it is out of scope of the survey) and -8 (the person was out-of-scope last wave already).

2.1.2. Weighting status of a person in the previous wave (variable name: PWSTAT_PREC)

This information is available from the link-file of the ECHP UDB. The variable PWSTAT_PREC copies the values of PWSTAT if these are > 0, otherwise, that is if PWSTAT < 0, the PWSTAT information from the most recent wave where PWSTAT is positive is used.

2.1.3. Sampling status of a person (variable name: SAMPERS)

This information is available from the link-file of the ECHP UDB. The variable SAMPERS has values 1 (sample person) or 2 (not a sample person).

2.1.4. Membership status of a person in the previous wave (variable name: PLASTW)

This information is available from the link-file of the ECHP UDB. The variable PLASTWi has value 6 for new-borns.

2.1.5. Eligibility of a person (variable name: PELIG)

This information is available from the link-file of the ECHP UDB. The variable PELIG has values 1 (the person is eligible for the personal interview) or 2 (the person is not eligible for the personal interview).

2.1.6. Individual interview result (variable name: PFNRES)

This information is available from the link-file of the ECHP UDB. The variable PFNRES has values 11 or 13 for interviewed persons (11 = full questionnaire, 13 = reduced questionnaire), any other value refers to persons for whom no answer to the questionnaire is available (this includes code 12 = interview completed, but not transmitted to Eurostat).

2.1.7. Age of individuals (variable name: AGE)

This information is derived from the linkfile of the ECHP UDB as follows:

	AGE
BIRTHYY ≠ -9	1993 + wave - BIRTHYY
else	.

2.1.8. Sex of individuals (variable name: SEX)

The sex is available from the linkfile of the ECHP UDB. Additionally, a binary variable is created:

	SEX1
SEX = 1	1
else	.

2.1.9. Category age-sex of individuals (variable name: AGESEX)

Based on their sex and age (see above), individuals are grouped into 20 categories

		AGESEX
SEX = 2	0 ≤ AGE < 16	1
	16 ≤ AGE < 20	2
	20 ≤ AGE < 25	3
	25 ≤ AGE < 35	4
	35 ≤ AGE < 45	5
	45 ≤ AGE < 55	6
	55 ≤ AGE < 60	7
	60 ≤ AGE < 65	8
	65 ≤ AGE < 75	9
	75 ≤ AGE	10
SEX = 1	0 ≤ AGE < 16	11
	16 ≤ AGE < 20	12
	20 ≤ AGE < 25	13
	25 ≤ AGE < 35	14
	35 ≤ AGE < 45	15
	45 ≤ AGE < 55	16
	55 ≤ AGE < 60	17
	60 ≤ AGE < 65	18
	65 ≤ AGE < 75	19
	75 ≤ AGE	20

2.1.10. Design weight of a household (variable name: DESIGNWT)

The design weight is inversely proportional to the sample selection probability. The sample selection probability for each household is known from the sample design.

The design weight is available from the linkfile of the ECHP UDB.

2.1.10.1. Estimate design weights when they are missing (DESIGNWT)

Missing design weights of households will be estimated as the average of design weights of households in the same region and of the same size.

2.1.11. Household size (variable name: HSIZE)

This information is copied from the H-file of the ECHP UDB.

HSIZE = HD001

Five binary variables are defined

HSIZE1

HSIZE = 1	1
else	.

HSIZE2

HSIZE = 2	1
else	.

HSIZE3

HSIZE = 3	1
else	.

HSIZE4

HSIZE = 4	1
else	.

HSIZE5

HSIZE > 4	1
else	.

2.1.12. Number of economically active persons in the household (variable name: NBACT)

This information is derived from the H-file of the ECHP UDB.

NBACT = HD013(from current wave), if not missing
HD013(from an earlier wave), if missing for current wave.

Three binary variables are defined

NBACT1

NBACT = 0	1
else	.

NBACT2

NBACT = 1	1
else	.

NBACT3

NBACT > 1	1
else	.

2.1.13. Arrivals to or departures from the household (variable name: DEP_ARR)

Based on information from the linkfile of the ECHP UDB, this variable can be derived.

DEP_ARR = HMIN + HBORN - HDIED - HMOUT

Three binary variables are defined:

DEP_ARR1

DEP_ARR < 0	1
else	.

DEP_ARR2

DEP_ARR = 0	1
else	.

DEP_ARR3

DEP_ARR > 0	1
else	.

2.1.14. Type of household (variable name: HHTYPE)

This information is derived from the H-file of the ECHP UDB.

HHTYPE = HD006 (from current wave), if not missing
HD006 (from an earlier wave), if missing for current wave.

Twelve binary variables are derived:

HHTYPE1

HHTYPE = 1	1
else	.

HHTYPE2

HHTYPE = 2	1
else	.

HHTYPE3

HHTYPE = 3	1
else	.

HHTYPE4

HHTYPE = 4	1
else	.

HHTYPE5

HHTYPE = 5	1
else	.

HHTYPE6

HHTYPE = 6	1
else	.

HHTYPE7

HHTYPE = 7	1
else	.

HHTYPE8

HHTYPE = 8	1
else	.

HHTYPE9

HHTYPE = 9	1
else	.

HHTYPE10

HHTYPE = 10	1
else	.

HHTYPE11

HHTYPE = 11	1
else	.

HHTYPE12

HHTYPE = 12	1
else	.

2.1.15. Tenure status (variable name: TENURE)

This information is derived from the H-file of the ECHP UDB.

TENURE = HD023 (from current wave)

This variable is transformed into a binary variable:

	TENURE1
TENURE = 1	1
else	.

2.1.16. Main source of income (variable name: INCMA)

This information is derived from the H-file of the ECHP PDB.

INCMA = H0i0830 , if H0i0830 > 0
H0(i-1)0830, else

Eight binary variables are derived

	INCMA1
INCMA = 0	1
else	.

	INCMA2
INCMA = 1	1
else	.

	INCMA3
INCMA = 2	1
else	.

	INCMA4
INCMA = 3	1
else	.

	INCMA5
INCMA = 4	1
else	.

	INCMA6
INCMA = 5	1
else	.

	INCMA7
INCMA = 6	1
else	.

	INCMA8
INCMA = 7	1
else	.

2.1.17. Split-off household (variable name: SPLITT)

This binary information is derived from the H-file of the ECHP UDB.

	SPLITT
HSTATUS = 2	1
else	.

2.1.18. Equivalised income (variable name: INC_EQ)

The equivalised income of a household is calculated as follows (from the H-file of the ECHP UDB):

INC_EQ = HI200 / HD005

2.2. Starting weight (WSTART_SP)

For each sample person, a positive starting weight is needed in the weighting procedure.

2.2.1. Initial wave

For the first wave, the starting weight for each individual is the design weight (see above). All the individuals present in the first wave are sample persons and, consequently, have a starting weight greater than 0.

2.2.2. Subsequent wave

For any subsequent wave, the starting weight for each individual is its final base weight from the previous wave (for sample persons this weight is > 0, for non-sample persons this weight equals 0).

2.3. Calculate probability of re-enumeration and of response (P1, P2 and P3)

The probability (P1) for an individual being resident in wave i - i.e. being member of an interviewed household -, if it was resident in a previous wave, is calculated as follows

P1 = P (PWSTAT = 1 | PWSTAT_PREC = 1)

The probability (P2) for an individual for having been resident in the last wave (i-1), if it is resident in the current wave is calculated as follows:

P2 = P (PWSTAT_PREC = 1 | PWSTAT = 1 and PLASTW ≠ 6)

The probability (P3) of being interviewed, if the person is eligible in the current wave.

P3 =
P (PFNRES(i) = 11,13 | PELIG(i)=1 and PWSTAT(i)=1 and SAMPERS=1)

In order to calculate these 3 probabilities, a logistic regression selects first the explanatory variables. Explanatory variables are selected from the following list:

Binary variables:

- REGION1…REGIONx (x=number of regions, see annexe 1 for regions)
- SPLITT
- DEP_ARR1 … DEP_ARR3
- INCMA1 … INCMA8
- NBACT1 … NBACT3
- HSIZE1 … HSIZE5
- SEX1
- TENURE1

Continuous variables:

- AGE
- INC_EQ

The probabilities are calculated with the SAS procedure CATMOD which models categorical data and fits linear models to functions of response frequencies. In order to avoid extreme weights, these probabilites are trimmed. P1 and P2 are trimmed to the 5th percentile, while P3 is trimmed to the 1st percentile.

2.4. Calculate provisional weights for all persons (WPROV_P)

For each sample person, the provisional weight is calculated by multiplying the starting weight (WSTART_SP) by P2 and dividing by P1.

WPROV_P = WSTART_SP * P2 / P1

For those sample persons that are member of an household that was added to the sample in the current wave, the provisional weight will be set to 1.

WPROV_P = 1.

The provisional weight of non-sample persons is 0.

WPROV_P = 0

New born children are assigned the provisional weight of their mother.

2.4.1. Estimate missing provisional weights for persons (WPROV_P)

If the provisional weight is missing for a sample person, it is estimated as a transformation of its design weight. Therefore, a polynomial regression is solved - by region, household size and number of economically active persons in the household.

$WPROV\_P = a * DESIGNWT^2 + b * DESIGNWT + c$

Remark: For countries with DESIGNWT = 1, the regression simplifies into

WPROV_P = b * DESIGNWT + c

2.5. Calculate provisional weights for households (WPROV_H)

For each household, the average of the provisional weights of all the sample persons in this household is calculated.

WPROV_H = Σ WPROV_P / number of sample persons in the household

2.6. Calibrate household weights (WCAL_H)

The provisional weights WPROV_H are calibrated in order to reflect the distribution of the population. The unit for which the calibration is executed is the household. However, the individual is taken into account as well. Variables concerning both levels are taken into account during the calibration.

The sample of households is to have the same structure as the total population of households by region (for breakdown see Annex I). For seven countries[2], a calibration by household size was applied as well (households are grouped into the categories '1 person household', '2 or more person household').

The sample of individuals is to have the same structure as the population for the following variables, i.e. for the cross-classification of age and sex:

- females aged 0 to 15
- females aged 16 to 19
- females aged 20 to 24
- females aged 25 to 34
- females aged 35 to 44
- females aged 45 to 54
- females aged 55 to 59
- females aged 60 to 64
- females aged 65 to 74
- females aged 75 or more
- males aged 0 to 15
- males aged 16 to 19
- males aged 20 to 24
- males aged 25 to 34
- males aged 35 to 44
- males aged 45 to 54
- males aged 55 to 59
- males aged 60 to 64
- males aged 65 to 74
- males aged 75 or more

The CALMAR software calibrates weights for households and individuals at the same time, by using one single microdata-file with one record per household. Therefore it is necessary to calculate, for each household, the number of sample persons belonging to each age-sex group. These variables are taken into account as numerical variables, while the variables at household level (tenure status, household size, …) are used as classification variables.

[2] The countries concerned are Germany, Ireland, Luxembourg, the Netherland, Austria, Finland and Spain. For Spain, however, three category were used '1 person household', '2 person household', '3 or more person household'.

In order to calibrate according to marginal distributions, a datafile describing the margins has to be available. This datafile contains the distribution of the the population of households and persons.

The logit method (method 3) of CALMAR is used for calibration. In order to define parameters needed for the logit method (LO, UP parameters) the raking method is used. In case of non-convergence, the parameters LO and UP are adjusted manually.

2.7. Calculate base weights of sample persons (RG003)

For each household, the calibrated household weight (WCAL_H) is assigned to each sample person in the household. Value 0 is assigned to non-sample persons. These weights are scaled such that their sum over all persons in interviewed households equals the actual number of persons in these households. This means that the average of base weights of sample and non-sample persons is 1.

The base weight of non-sample persons is 0.

2.8. Calculate cross-sectional weights of persons (RG002)

All residents (including sample and non-sample persons) of an interviewed household receive the same cross-sectional weight, computed as the average of base weights (RG003) of the household members. This means that the sum of cross-sectional weights of persons in a household equals the sum of their base weights, which also implies that for the whole sample the cross-sectional weights are scaled such that their sum equals the total number of residents in households, i.e. the average per person is 1.

The assignment of non-zero weights to all residents permits the inclusion of non-sample persons in cross-sectional analysis.

2.9. Calculate cross-sectional weights (HG004)

Each household is assigned the cross-sectional weight of the household members. This means that the cross-sectional household weight is proportional to the cross-sectional weights of persons. The houehold weights are scaled such that their sum over the interviewed sample equals the actual number of completed houdehold interviews. This means that the average household weight is 1.

2.10. Calibrate weights of interviewed sample persons (WCAL_P)

The calibrated household weight WCAL_H is assigned to each interviewed sample person and divided by P3 (the probability of being interviewed, when eligible).

2.11. Calculate base weights of interviewed persons (PG003)

The calibrated weights (WCAL_P) are available for each interviewed sample person. Value 0 is assigned to non-sample persons. These weights are scaled such that their sum over the interviewed persons equals the actual number of

interviewed persons. This means that the average of base weights of interviewed sample and non-sample persons is 1.

The base weight of interviewed non-sample persons is 0.

2.12. Calculate cross-sectional weights of interviewed persons (PG002)

All interviewed persons (including sample and non-sample persons) receive the same cross-sectional weight, computed as the average of base weights (PG003) of all interviewed household members. This means that the sum of cross-sectional weights of persons in a household equals the sum of their base weights, which also implies that for the whole sample the cross-sectional weights are scaled such that their sum equals the total number of interviewed persons in households, i.e. the average per person is 1.

The assignment of non-zero weights to all interviewed persons permits the inclusion of non-sample persons in cross-sectional analysis.

3. ANNEX 1 - REGIONS

Germany
1="Baden-Württemberg"
2="Bayern"
3="Berlin"
4="Brandenburg"
5="Bremen"
6="Hamburg"
7="Hessen"
8="Mecklenburg-Vorpommern"
9="Niedersachsen"
10="Nordrhein-Westfalen"
11="Rheinland-Pfalz"
12="Saarland"
13="Sachsen"
14="Sachsen-Anhalt"
15="Schleswig-Holstein"
16="Thüringen"

Danmark
1="Koebenhavn + Frederiksberg"
2="Koebenhavn AMT"
3="Frederiksborg AMT"
4="Roskilde AMT"
5="Vestsjaellands AMT"
6="Storstroems AMT"
7="Bornholms AMT"
8="Fyns AMT"
9="Soenderjyllands AMT"
10="Ribe AMT"
11="Vejle AMT"
12="Ringkoebing AMT"
13="Aerhus AMT"
14="Viborg AMT"
15="Nordjyllands AMT"

Netherlands
1="Noord-Nederland"
2="Oost-Nederland"
3="West-Nederland"
4="Zuid-Nederland"

Belgium
1="Bruxelles-Brussels"
2="Vlaams Gewest"
3="Region wallone"

France
1="Ile de France"
2="Bassin Parisien"
3="Nord - Pas-de-Calais"
4="Est"
5="Ouest"
6="Sud-Ouest"
7="Centre-Est"
8="Méditerranée"

United-Kingdom
1 = "North"
2 = "Yorks & Humberside"
3 = "North-West"
4 = "East Midlands"
5 = "West Midlands"
6 = "East Anglia"
7 = "London"
8 = "Other South-East"
9 = "South-West"
10 = "Wales"
11 = "Scotland & Northern Ireland"

Ireland
1="East"
2="South-West"
3="South-East"
4="North-West"
5="Mid-West"
6="Donegal"
7="Midlands"
8="West"
9="North-West"

Italy
1="Nord Ovest"
2="Lombardia"
3="Nord-Est"
4="Emilia-Romagna"
5="Centro"
6="Lazio"
7="Abruzzo-Molise"
8="Campagia"
9="Sud"
10="Sicilia"
11="Sardegna"

```
Greece
   1="Voreia Ellada"
   2="Kentriki Ellada"
   3="Attiki"
   4="Nisia Aigaiou, Kriti"

Spain
   1="Galicia"
   2="Principado de Asturias"
   3="Cantabria"
   4="Pais Vasco"
   5="Communidad Foral de Navarra"
   6="La Rioja"
   7="Aragon"
   8="Communidad de Madrid"
   9="Centro"
   10="Castilla-la Mancha"
   11="Extremedura"
   12="Cataluna"
   13="Communidad Valenciana"
   14="Islas Beleras"
   15="Andalucia"
   16="Region de Murcia"
   17="Canarias" ;

Portugal
   1="Açores"
   2="Madeira"
   3="Norte"
   4="Centro"
   5="Lisboa e Vale do Tejo"
   6="Alentejo"
   7="Algarve"

Austria
   1="Ostösterreich"
   2="Südösterreich"
   3="Westösterreich"

Finland
   1="Manner-Suomi"
   2="Etela-Suomi"
   3="Ita-Suomi"
   4="Vali-Suomi"
   5="Pohjois-Suomi"
```