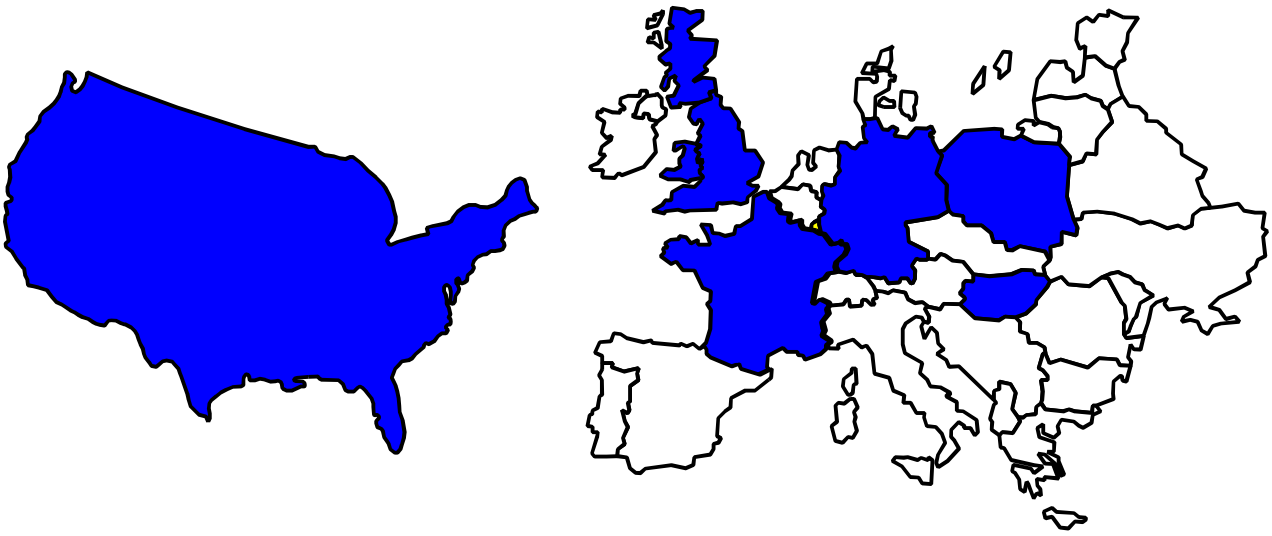
PACO
panel
comparability



User Manual

Version 30.9.97

(http://www.ceps.lu)
(e-mail: gunther.schmaus@ci.rech.lu)

Comparative Research on Household Panel Studies

PACO

Document n° 9 1997

PACO USER GUIDE

by

Günther Schmaus
Marlis Riebschläger

***

This publication was supported by the Human Capital and Mobility Programme, Directorate General for Science, Research and Development of the Commission of the European Communities.

***

Comparative Research on Household Panel Studies

This series presents the results of research projects based on the analysis of one or more household panel studies. Papers will cover the wide range of substantive topics and investigations of the particular problems of comparative research.

The series will contain, among other papers, the results of all of the work being carried out as part of the Panel Comparability (PACO) project, which was funded by the European Commission under the Human Capital and Mobility Programme (1993-1996). PACO aims to develop instruments for analyzing, programming and stimulating socio-economic policies, and for comparative research on policy issues such as labour force participation, income distribution, unpaid work, poverty, household composition change, and problems of the elderly.

Coordination of the project is provided by

CEPS/INSTEAD, Differdange, Luxembourg.

Associated partners are:

- German Socio-economic Panel Study (SOEP), Deutsches Institut für Wirtschaftsforschung (DIW) Berlin
- British Household Panel Study (BHPS), ESRC Research Center, University of Essex
- Lorraine Panel Study, ADEPS/URA Emploi et Politiques Sociales, Nancy
- Economic and Social Research Institute (ESRI), Dublin
- Gabinet d'Estudis Socials (GES), Barcelone
- Luxembourg Household Panel Study (PSELL), CEPS/INSTEAD Differdange

- Hungarian Household Panel (HHP); TARKI Budapest
- University of Warsaw, Dept. of Economics, Warsaw
- Institute of Sociology, Academy of Sciences of the Czech Republic, Prague

Associated projects are the Female Labour Force Participation Project, also funded under the European Commission Human Capital and Mobility Programme, and the Network of Host Centres on Comparative Analysis of European Social Policy, as well as other research based on household panels.

The editing of this series was done under the guidance of Marcia Taylor, PACO network coordinator at CEPS/INSTEAD (1993-1996).

For more information about this series, or to submit papers for inclusion, contact:

CEPS/INSTEAD
Anc. Bât. administratif ARBED
Rue E. Mark, Boîte postale 48
L- 4501 Differdange
Tel: +352 58 58 55-555
Fax: +352 58 55 88

Document n° 9: ***PACO USER GUIDE;***
by Günther Schmaus, Marlis Riebschläger.
CEPS/INSTEAD, Luxembourg, 1997.

 ISBN 2-87987-085-2

Table of Contents

A. Introduction and Overview on Panel Studies Included in the PACO Database

The aim of the PACO (**PA**nel **CO**mparability) project is the creation of a harmonized and standardized micro-database from existing longitudinal studies on living conditions of households.

The project team involved researchers from France, Germany, Hungary, Ireland, the United Kingdom, Luxembourg, Poland and Spain. The coordination was held at the CEPS institute in Luxembourg.

The project was partly funded by the European Commission from 1993 to 1995

Currently the following panel studies are included:

BHPS : British Household Panel Study
(1991 - 1993, ongoing)
ESEML : Enquête Socio-Economique des Ménages en Lorraine / France
(1985 - 1990, study ended in 1990)
HHP: Hungarian Household Panel Study
(1992-1994,ongoing)
PHP: Polish Household Panel Study
(1987-1990,ongoing)
PSELL : Panel Socio-Economique Liewen zu Lëtzebuerg / Luxembourg
(1985 - 1992, ongoing)
PSID : Panel Study of Income Dynamics / USA
(1983 - 1987; the panel was started in 1968, but in order to cover more or less comparable ranges of years, only waves from 1983 onwards were included.)
SOEP : Sozio-Oekonomisches Panel / Bundesrepublik Deutschland
(1984 - 1991, ongoing)

Further waves will be added in future.

The Starting Point

Without a harmonized database cross-national comparative studies on panel data are feasible only by teams involving members of the domestic panel staffs. This is due to the complexity and to the differences in the organization of the panel databases.

The main differences can be summarized as follows:

- Questions concerning the same topics are asked in different manners. Even if the questions are comparable, different categories are built. Even standard demographic variables are coded in different ways. Missing values are also coded differently.
- The levels on which information is collected differs. In one panel study a question might be asked on the individual level whereas in another study a similar question is asked on the household level. In the Luxembourg and Lorraine panel there is a third level - the income group: In a household in which several persons have individual income, different economic arrangements are possible. An income group is a group of persons within a household, who constitute an economic unit. In the PSID most information on the individual level is collected only for the head of the household and the spouse.
- Different storage formats or database systems are used:
  ASCII (PSID), SIR (SOEP, BHPS), SAS (SOEP public use version), SPSS(PSELL), SYBASE (ESEML)
- The files are structured differently:
  The most simple but also space consuming structure is set up by two longitudinal files, one comprising all the households and the other comprising all individuals who ever took part in the panel (PSID).
  The PSELL is stored in three files per wave, one file for households, the second for individuals and the third file is set up by the income groups
  The SOEP and BHPS have a more complex structure. On both the household and individual level the information which is obtained directly from answers to questions in the questionnaire is separated from fieldwork information and also from additional generated information, i.e. updated information on questions which are asked only if a change has occurred. The data on children up to the age of 15, who are not interviewed, but on whom information is collected via the household questionnaire, are also stored in an extra file. Additional longitudinal files for households and individuals are supplied containing one record per household or individual, respectively, by which the trajectories can be followed. Files containing job history data and biography data are also supplied.

- The naming conventions are different in the different panels:
  The PSID simply enumerates all the variables over all the years. In the PSELL system the variable names remain constant over the waves whenever the question has remained the same. Only a wave indicator is added. In the SOEP the names of variables which are related directly to questions are made up of the question number and a wave indicator, which means that they differ from wave to wave, since the order in the questionnaires does not remain the same. The names of fieldwork and generated variables are so-called "speaking" names, but they are speaking German. In the BHPS all the names are created to be

"speaking" (English speaking) but the length is limitated to 8 characters so the abbreviations are of limited value only.

In order to overcome the problems linked to the differences listed above, the PACO team has undertaken all efforts to standardize and harmonize the different data-bases in the following way:

Technical issues:

1. The Luxembourg way of data storage and naming conventions were adopted, which means:

- The variable names are constant except for a year indicator.
  The first character indicates the level of information : (P = Person, G = Income Group (Luxembourg and Lorraine), H = Household
- The data are stored as SPSS-files. For each country and each year one file for individuals (including children), and one file for households was created. For Luxembourg and Lorraine a file for income groups was created for each year as well. The file names contain information on the year, the country and the level of information (household, income group, or individual).
  In addition to that there are cross-year files containing time independent information.

2. Variables were recoded or categories were regrouped in order to obtain common variables for the different countries. This includes also the recoding of missing values.
3. The information is supplied on the lowest possible level, which means the most detailed level.
   The income variables are also aggregated to higher levels (income group, household), in order to supply comparable variables when the level of information was different in the different countries.

Ideally the analysis can be performed as follows:
Once a program for data analysis has been written for one country, it can be run for the other countries in the same way (if all the information is available in all the countries; see explainations below).

Contents

The following topics are covered:

The income variables are very detailed (there are 66 of these). The list was set up in order not to loose any information which is available in at least two countries. For the other topics, the selection is less detailed. One can say that in general a compromise was made between not loosing too much information on the one hand and not keeping too much detailed information and ending up in keeping variables which are available for one country only, on the other hand.

The PACO team tried to use standardized coding schemes whenever these were available. For example the ISIC and ISCO codes were applied to the employment sector and the type of occupation, respectively.

The OECD classification was used for the education variables because this scheme covers the USA as well. Since this scheme is very global, an additional variable was included to keep more details for countries to which they apply.

In addition to the variables which have been extracted from the original panel data, other variables were added to the dataset.

A variable which provides a tool to apply the Random Group or Jackknife Method for variance estimation was generated for each country.

Furthermore a set of macro-economic indicators is provided together with variables that indicate the country and the year, so that these variables can be linked easily to the micro-data.

Access to PACO

The PACO Database is adequately anonymized and accessible as a set of scientific use files under appropriate conditions regarding confidentiality and data protection. It is available on a CD-ROM, together with data documentation, a set of macro-variables and relevant parts of the MISSOC publications on Social Security.

For detailed information about the PACO project and research papers, please contact:

CEPS/INSTEAD

Anc. Bâtiment Administratif 1 ARBED
Rue Emile Mark ,B.P. 48
L - 4501 Differdange
Tel: (00352) 58 58 55 -1
Fax: (00352) 58 55 88

E-mail:

gaston.schaber@ci.rech.lu
gunther.schmaus@ci.rech.lu

Internet: http://www.ceps.lu/

B PACO Database Definition

B.1 General Remarks

Introduction

The PACO DATA BASE contains harmonized and consistent variables and identical data structures for each country included. It can therefore increase the accessibility and use of panel data for research. **The PACO DATA BASE can be used to ease comparative cross-national and longitudinal research and to study processes and dynamics of policy issues such as labour force participation, income distribution, poverty, problems of the elderly** and so on.

PACO DATA BASE

PACO adds value to the original panel data by creating **Compatibility** and **Comparability**. The process of making data comparable is realized by creating harmonized and consistent variables and files. **The PACO Data Base contains comparable variables transformed according to a common plan and was built by using standardized international classifications where available. Information in these files is available (a) for households and individuals on the micro level, (b) for single years and (c) as longitudinal information.** Such a comparative approach results in a (common) PACO Data Base, currently containing the data from seven countries (additional countries being included later). All files are held in a relational Data Base Structure. The data are stored as system files for the statistical package SPSS for Windows: containing **identical variable names, labels, values and data structures**. Each country file is adequately anonymized and can therefore be rated as a public use file. The complete data base has a size of 250 MB and is available on CD-Rom.

Advantages of PACO

The PACO approach - using highly standardarized variables and files - facilitates the analysis of cross-national panel data: a) Macro utilities enable the user to retrieve and to match the PACO data more easily. b) The PACO data structure allows to write global analyses programs. c) **Standard analyses programs can be run for different countries and different periods with no need to modify the underlying SPSS (Macro) programs.** d) **The processing of PACO files is easier than analysing the original panel studies.** e) The researcher has not to be familiarized with the individual panel's data organization.

Moreover, the PACO user has the possibility of processing those original variables in the panel studies that have not been made comparable. The researcher can simultaneously access original variables from original panel studies and the harmonized variables from the PACO DATA BASE.

Documentation

Each PACO variable is fully documented including information on the algorithms used in variable creation and an indication of the comparative reliability of each variable. This documentation can be find in the national variable documentation (not included in this PACO User guide).

The PACO DATA BASE can be linked with a collection of macro data. A set of macro variables were extracted from the EUROSTAT CD of year 1993 and other statistical sources. The macro data is accessible from SPSS and can be matched with the PACO files.

The relevant parts of the MISSOC publications about Social Security have been compiled and integrated into the PACO Documentation system. The available information allows to link original variables from national panel studies with the MISSOC data; on the other hand it is possible to retrieve the MISSOC information about selected PACO variables. The MISSOC-PACO link enables the interpretation of results from cross-national research with the PACO DATA BASE.

B. 2 Definition of Units

a) Definition of Households

A household consists of all persons who live together in a dwelling unit (house, appartment, group of rooms or single room). Persons within a household can be related to each other or not. Included are unmarried couples, if the couple is living in a fairly permanent arrangement. A household may consists of more than one family, if the persons are related to each other. Excluded are lodgers, conventional roommates, or employees who share the housing unit.

b) Definition of Income groups

In a household in which several persons have individual income, different economic arrangements are possible. An Income group is a group of persons within a household who constitute an economic unit because they share their incomes. The concept is put into effect in accordance with strict rules and according to responses from household members in the interview.

examples:

- If a household consists of a couple with minor children **without** own incomes than there is only **one** income group in the household, because it is a priori assumed that partners in a household share their incomes.

- if a household consists of a couple with one adult children **with** own income than there may be **two** income groups: The first income group is assembled by the parners of the couple, the second income group by the adult children.

The concept of Income groups is originally only used within the French(Lorraine, ESEML) panel and the Luxembourg panel (PSELL).

Due to the data strucure of the American PSID files some income variables are only available as a sum for head and spouse together, other income variables are available only as a sum for all other family members (not head or spouse). For those type of variables two artifical income groups have been created.

B.3 Variable Specifications for PACO Database

B.3.1. List of variables

B.3.1.1. Income variables

_xx001	***Total wages and salaries***
_xx002	Wages and Salaries
_xx003	Salary from apprenticeship
_xx004	Salary from working student
_xx005	Second Salary
_xx006	Employer Bonuses for employees
_xx007	In kind Salary
_xx008	***Total self-employment income***
_xx009	Self-employment Income from enterprises
_xx010	Income from independent professionals
_xx011	Farm self-employment Income
_xx012	***Total earnings***
_xx013	***Total property income***
_xx014	Income from interest and dividends
_xx015	Income from rents
_xx016	Income from owner occupied houses
_xx017	***Total factor income***
_xx018	Total employer pension
_xx019	Private pensions (Occupational pensions)
_xx020	Public Sector pensions
_xx021	***Total market income***
_xx022	***Total old age pension***
_xx023	Old Age Pensions
_xx024	Widow/Widower Pensions
_xx025	Orphan pensions
_xx026	***Total Retirement income***
_xx027	***Total social insurance income***
_xx028	Unemployment Benefits from Insurance
_xx029	Sickness Cash Benefits
_xx030	Employment Injuries/Occupational diseases benefits
_xx031	Invalidity Benefits
_xx032	War related benefits
_xx033	Family Benefits
_xx034	Maternity benefits
_xx035	Government cash transfers for Education
_xx036	***Total means-tested incomes***
_xx037	Social Assistance
_xx038	Additional Social Assistance
_xx039	Unemployment assistance
_xx040	Other Income Dependant Benefits
_xx041	***Total social security income***
_xx042	***Total private income***

_xx043 Cash Alimony or Child Support
_xx044 Received private Cash Inter-household Transfers
_xx045 ***Total transfers (excl old age pensions)***
_xx046 Transfers for Handicapped
_xx047 Other Transfers
_xx048 Annuities from insurance
_xx049 Other Lump-sum Payments
_xx050 Income from home consumption, or from farming and gardening
_xx051 Other Incomes
_xx052 ***Total transfers***
_xx053 ***Total gross income***
_xx054 ***Total social security contributions***
_xx055 Health Insurance
_xx056 Old Age Insurance
_xx057 Unemployment Insurance
_xx058 Other Direct Taxes
_xx059 Income tax
_xx060 ***Total net income***
_xx061 ***Total Contributed Private Cash Inter-household Transfers***
_xx062 Contributed Transfers to Parents
_xx063 Contributed Transfers to Children
_xx064 Contributed Transfers to Spouses
_xx065 Contributed Transfers to other Relatives
_xx066 Contributed Transfers to others

B.3.1.2. Demographic variables

P201 Sex
P202 Birth year
P203 Birth month
Pxx204 Family Status
Pxx205 Relationship to Reference Person in HH
Pxx206 Foreign Nationals
Pxx207 Degree of Handicap
Pxx208 Visits to Doctor (excluding dentists)
Pxx209 Subjective Health Status
Pxx210 Cohabitor Status
Hxx250 No. of Adults in HH
Hxx251 No. of Children in HH
Hxx252 Actual age of youngest co-resident child
Hxx253 Actual age of oldest co-resident child
Hxx254 No. of Adult Children in HH
Hxx255 No. of Non-resident children
Hxx256 Household Typology

B.3.1.3. Labour Force and Work history variables

Pxx301 Employment Status
Pxx302 Employment Type
Pxx303 Prof.Stat.(long version)
Pxx304 Prof. Stat.(short version)
Pxx305 Normal Working Hours Weekly
Pxx306 Total Working Hours
Pxx307 Overtime Compensation
Pxx308 Second Job
Pxx309 Working Hours Second Job
Pxx310 Employment Sector
Pxx311 Type of Employer
Pxx312 Firm Size
Pxx313 Type of Current Occupation
Pxx314 Contractual Situation
Pxx315 Total month in employment in most recent job
Pxx316 Employment status per year (retrospective)
Pxx317 Labour Force Status
Pxx318 Unemployment Status
Pxx319 Immediate Intention for Work
Pxx320 Future Intention for Work

B.3.1.4. Education variables

P401 Highest obtained School Education
P402 Highest obtained 2nd Lev. 2nd Stage Education
P403 Highest obtained Imputed Years of Education

Pxx401 Current School Education
Pxx402 Current 2.lev 2.stage Education
Pxx403 Current years of education

B.3.1.5. Housing variables

Hxx601 House Ownership Status

B.3.1.6. Other variables

Hxx801 Territorial division
Pxx810 Time spent on unpaid housekeeping work
Pxx811 Time spent on unpaid child care
Pxx812 Time spent on other unpaid care
Hxx813 Child care external to household

B.3.1.7. Weighting variables

PxxWEIG	Individual Weight
PxxWEIX	Individual Weight - special weight for Luxembourg to incl. the extension in 1991
PxxWEIL	Individual Longitudinal Weight
HxxWEIG	Household Weight
PxxPROB	Individual Retention Probability
HxxPROB	Household Retention Probability

B.3.1.8. Organizational (Link) variables

_L01	Country
_L02	Year
_xxL03	ID-Household
_xxL04	Pre-Year ID-HH
_xxL05	ID-Group
PL06	ID-Individual
_xxL07	ID-Reference Person
_xxL08	ID-Spouse of Reference Person
PxxL09	ID-Spouse
PL10	ID-Father
PL11	ID-Mother
_L12	Case ID
_L13	Random Group ID
_xxL14	Match Indicator

B.3.2. Convention of PACO Variable names

Year related variables: _xxttt
Non-year related variables: _ttt

where _ depends on the level of information which is available

This might be:

P = Person
G = Income Group (Lux and Lorraine)
H = Household

where **xx** denotes the year

85 = 19**85** 86 = 19**86** ... **xx** = 19**xx**

where **ttt** denotes a special variable domain:

001 to 199 Income variables
201 to 299 Demographic variables
301 to 399 Labour Force and Work history
401 to 499 Education and Family background
501 to 599 Marriage and Fertility
601 to 699 Housing
701 to 799 Summary of Calendar variables
801 to 999 other variables
L01 to L99 Organizational or Link Variables

other reserved first characters _

R = reference person ('head of household')
S = partner of reference person ('spouse')
C = partner ('Conjoint')
F = father
M = mother

B.3.3. PACO (User) Missing Codes

The missing codes for all variables with the exception of income variables are defined as follows:

-1 : Missing information
-2 : Inapplicable
-3 : Variable not available

B.3.3.1 Handling of Income variables with missing values

We assign zeros for each missing value for income variables. In addition to it we create one indicator variable which gives us the information about the status of that variable:

Value of the indicator variable for one income variable:

0 variable is not available :

- variable is not available for all units
- variable is not available for certain groups in the sample (e.g. persons other than head or spouse in the PSID.)

1 O.K.:

- This variable is available, this means a 0 entry is a valid value.

2 imputed values for missing values

3 missing

because of
- item non response
- unit non response

The value of the corresponding income variable is either missing or in the case of an aggregate variable there was at least one missing component.

	Income indicator	Income variable	Interpretation
a)	0	0	Variable is not available.
b)	1	0	valid entry
c)	1	...0	valid entry
d)	2	...0	imputed value(s)
e)	3	0	missing value(s)(see below)
f)	3	...0	

e) This could happen in the case of
- a single variable where the item was missing for the unit, or
- an aggregate variable, where at least some components are missing and the others have a valid zero entry.

f) This could only happen in the case of
an aggregate variable, where at least one component has a valid entry other than zero and the others are missing.

Possibilities of analysis

1. Possibility: If we select all records with value (1) we get all units with reported income.

2. Possibility: If we select all records with values (1) and (2) we get all units with reported income and also units which have imputed values.

3. Possibility: If we select all records with values (3) we get all units with missing values where the missing value was replaced by zero values.

Variables with value (0) cannot be analyzed (if this applies only to subgroups, this subgroups have to be eliminated from the analysis).

All combinations (0-3) are possible. Following rules are introduced to handle this problem:

It makes only sense to do summation over variables with values (1),(2),(3). For such units the following summation rule for indicator variables is used.

Summation Rules:

1. Rule: If at least one of the summation variables contains an imputed value (2) then the summation variable contains also imputed values.

2. Rule: If at least one of the summation variables contains a missing value (3) then the summation variable contains also missing values.

3. Rule: If the summation variables contain reported values (0), imputed values (2) and missing values (3) then the summation variable contains also missing values.

For aggregation it makes sense to aggregate individual incomes as above if the indicator variable contains values (1),(2) and (3). Units with value (0) which result from children records can also be aggregated, but aggregations of variables with value (0) which result from other cases are not allowed.

Aggregation Rules:

1. Rule: If at least one of the individuals in a household has a income with a imputed value (2) for one specific income variable then the aggregated variable contains also imputed values.

2. Rule: If at least one of the individuals in a household has a missing value (3) for one specific income variable then the aggregated variable contains also missing values.

3. Rule: If at least one of the individuals in the household has an income with imputed values (2) and another person in the household has a missing value (3) for one specific income variable then the aggregated variable also contains missing values.

4. Rule: Only variables with value (0) for children records may also be aggregated.

B.3.4. Definition of PACO Variables

B.3.4.1. Income Variables

B.3.4.1.1. Notes concerning Income variables

The following PACO Income list explains how country-specific income elements were assigned to comparable internationally consistent income categories.
As to increase comparability, the PACO Income list is widely consistent with the list of the income variables of the Luxembourg Income Study (LIS).
For those who are familiar with the LIS data the following hints might be useful.

The LIS income components list has been used as a basic model. For various reasons some modifications from the LIS list were necessary:

- some important income sources are much more detailed in PACO than in LIS
- LIS Government Non cash Transfers have been dropped
- LIS Variables with Employer Contributions for Social Security have been dropped
- some variables are slightly differently defined in PACO and LIS

Despite theses differences, The PACO income totals especially (for definition of totals see Part C, totals are marked in Italics) are identical with the LIS total variables.

Notes:

1) We have stored information both on the **household** and on the individual level. Where possible and feasible income variables were also created for the **group** level. First the information has been generated on the lowest level, e.g. the person-level if possible. In a second step the income of the person-level has been aggregated to the household level.

2) The income sources are in almost all cases **Gross incomes**.

3) We have supplied **cash income components**. The only exception are in kind salaries.

4) The income sources for transfers include only **positive cash transfers** and exclude tax allowances and in kind transfers.

5) The income variables are exclusively **continuous variables and income amounts;** negative values are allowed. 'Income brackets' have been converted into amounts.

6) The amounts are **monthly incomes** in the currency of the country.

a) If the original amount is a yearly value, the amount has been divided by 12.

b) If the number of months and the monthly amounts are known, the following formula has been used, to create monthly values:

PACO monthly amount = (no. of months x monthly amount)/12.

c) If only the last monthly amount is known or the reference period is unclear, we assume that it this is the right monthly value.

7) Some original variables may be not detailed enough to be assigned to PACO income elements. Whenever, if possible, **imputation rules** have been **developped** for creating PACO variables. Example: If only 'Total Self-Employment Income' is known but information about professional status(Entrepreneur, Liberal professions, Farmer) is available, it is possible to split 'Total Self-employment Income' into the components 'Self-employment Income from enterprises', 'Income from independent professionals' and 'Farm self-employment Income').

8) How to proceed with income variables which cannot be assigned correctly ?

a) First we have tried to allocate a specific variable to a variable which is most similar to the PACO variable.

b) Secondly, if this was not possible - because the specific variable is not detailed enough - we have assigned the original variable to a 'total' variable, which is most similar to the original variable.

c) Only if both mentioned procedures were not feasible, we used the variables 'Other transfers' and 'other incomes'.

9) Difference between Private Pensions/Public Sector Pensions/Old Age Pensions:

Private Pensions are paid by the **employer** (or via his/her insurance) only as a supplement to the Old Age Pensions of the State system. Private pensions are based on voluntary agreements between employer and employees and are not state regulated.

In some countries some employees of the state (not civil servants) receive additional Occupational Pensions as a supplement to the Old Age Pension system. These pensions were be counted as private pensions.

Public Sector Pensions are paid by the **state** for **its employees** (civil servants). Civil servants receive only Public Sector pensions and no further Old Age Pensions. The regulations about Public Sector Pensions are not based on voluntary agreements between the state and individual employees.

Old Age Pensions are paid from the **regular state pension system** to all employees (and former self-employed), who have paid contributions to the pension system. The membership in the state pension system is mostly compulsory and the amount of contribution is determined by social legislation.

b) Invalidity benefits are state benefits for disability of former employees which have paid contributions to the old age pension system. Individuals receiving disability need necessarily not be handicapped.

Transfers for handicapped (e.g. the blind) are to be paid to all individuals.

10) Difference between Wages and Salaries and Transfers

Benefits (for sickness, maternity, pre-retirement, etc.) which are paid directly by the employer are counted as Wages and Salaries and have not been allocated to the different social security benefit categories.

11) Difference between means-tested and not means-tested benefits

All benefits which are not means-tested have been allocated to the different appropriate social security categories.

12) Differences between
_xx030 (Employment Injuries/Occupational diseases benefits),
_xx031 (Invalidity Benefits) and
_xx046 (Transfers for Handicapped):

a) _xx030:

The benefits from insurance for Employment Injuries/Occupational diseases are paid to compensate for health problems which are related to injuries which had happened at work and also for occupational diseases which are related to the working situation. Therefore only employed employees (in some countries also employers) can receive this benefit. In most countries the premium for this insurance is exclusively paid by the employer.

b) _xx031:

Invalidity benefits are paid for employees who - as a result of sickness or infirmity - have lost the working capacity to a certain degree. Therefore invalidity benefits are typically paid for those

cases where the sickness or infirmity is not related to employment injuries or occupational diseases. Invalid individuals receive a pension from the state old age pension insurance.

Under certain conditions employees with a benefit from employment injury may receive in addition to it invalidity benefits also.

c) _xx046:

The transfers for handicapped include all benefits for seriously handicapped persons which are not listed under _xx030 and _x0031. The most relevant difference to _xx030 and _xx031 can be seen in the fact that also individuals without any labour market activities (e.g. housewifes with handicapps and parents for their handicapped child) could receive this type of benefit.

Transfers for handicapped are typically state transfers and are not insurance based.

General remark on aggregate variables :

The definition of the aggregate variables (see e.g. _xx001) is an ideal one. The value of an aggregate variable is not necessarily the sum of its components as is described in the definitions.
When for example smaller components are missing, the total was still calculated. In the Luxembourg data neither gross-income nor taxes are available; hence the net-income was assigned directly to the PACO-variable instead of using the formula in this definition.

B.3.4.1.2. Income variables definitions

_xx001 TOTAL WAGES AND SALARIES

=	_xx002	Wages and Salaries
plus	_xx003	Salary from apprenticeship
plus	_xx004	Salary from working student
plus	_xx005	Second Salary
plus	_xx006	Employer Bonuses for employees

_xx002 Wages and Salaries

only income from non self-employment

includes:

- normal wages and salaries
- premium for piece-work, incentive pay
- commissions
- overtime payment
- premium for night and weekend work
- family and child allowances paid by employer
- sick pay from employer (statutory sick pay)
- statutory maternity pay from employer
- pre-retirement income from employer
- allowances for partial unemployment paid by employer
- indemnity from employer for discharge of staff (redundancy pay)
- employee contributions for social security

excludes:

- employer's contributions for payroll (social insurance) taxes
- second salary

_xx003 Salary from apprenticeship

includes also on-the-job training

_xx004 Salary from working student

_xx005 Second Salary

income from secondary non-self-employment activity

_xx006 Employer Bonuses for employees

13 th and 14 th salary, one off bonus (e.g. at the end of the year), holiday bonus ("Urlaubsgeld"), other bonuses not paid monthly

_xx007 In kind Salary

fringe benefits (product sharing, remuneration in kind)

_xx008 TOTAL SELF-EMPLOYMENT INCOME

=	_xx009	Self-employment Income from enterprises
plus	_xx010	Income from independent professionals
plus	_xx011	Farm self-employment Income

_xx009 Self-Employment Income from enterprises

includes:

- distributed profits
- undistributed profits
- income from free-lancers
- social insurance contributions

excludes:

- profits from shares
- income of farmers
- capital gains
- income from independent professionals

_xx010 Income from independent professionals

Income from independent professions (as such physician,lawyer, tax-expert, engineer, architects etc.)
includes social insurance contributions

Individuals receiving this income source must have a university degree or higher, must be self-employed and executing a job listed above.

_xx011 Farm Self-employment Income

includes:

- money income from farming
- income in kind (money equivalent)
- social insurance contributions

excludes: income from farming and gardening of individuals with main occupations other than farmer. Their incomes are counted under _xx050.

_xx012 TOTAL EARNINGS

=	_xx001	Total wages and salaries
plus	_xx008	Total self-employment income

_xx013 TOTAL PROPERTY INCOME

=	_xx014	Income from interest and dividends
plus	_xx015	Income from rents
plus	_xx016	Income from owner occupied houses

_xx014 Income from interest and dividends

includes:

- interest from savings
- interest from building and loan association
- interest from life insurance
- interest from shares and loans
- income from trust funds
- income from royalties

_xx015 Income from rents

- Income from renting out houses and estates
- Income from lodgers and boarders

_xx016 Income from owner occupied houses

Following formula could be used to calculate income from owner occupied houses:

estimated rental value of house
minus interests for mortgages
minus costs for utilities
minus costs for heating
minus maintenance costs
= income from owner occupied houses

include also value of free housing for those who neither own or rent

_xx017 TOTAL FACTOR INCOME

= _xx012 Total earnings
plus _xx013 Total property income

_xx018 TOTAL EMPLOYER PENSION

= _xx019 Private pensions (Occupational pensions)
plus _xx020 Public Sector pensions

_xx019 Private pensions (Occupational pensions)

includes all supplementary pension schemes from (private)employers. The pension may also be paid via private insurance if the employer has paid the premium.
Includes also benefits for survivors from occupational pensions.

include also supplementary old age pensions for workers and employees (not civil servants) in public service

excludes pensions from private insurance if the employee has exclusively paid the premium

_xx020 Public Sector pensions

includes all old age pensions which are paid directly from a state employer to its civil servants:

Includes:

- old age pensions for civil servants
- Includes also benefits for survivors.

_xx021 TOTAL MARKET INCOME

= _xx017 Total factor income
plus _xx018 Total employer pension

_xx022 TOTAL OLD AGE PENSION

= _xx023 Old Age Pensions
plus _xx024 Widow/Widower Pensions
plus _xx025 Orphan pensions

_xx023 Old Age Pensions

include

- only pension from the state pension system for individuals with own entitlement
- pre-retirement pensions from the state

exclude: Public Sector pensions
Private Pensions
Widow/Widower pensions
Orphan pensions

_xx024 Widow/Widower Pensions

include here only pensions from the old age pension system

_xx025 Orphan pensions

include here only pensions from the old age pension system

_xx026 TOTAL RETIREMENT INCOME

= _xx018 Total employer pension
plus _xx022 Total old age pension

_xx027 TOTAL SOCIAL INSURANCE INCOME

=	_xx022	Total old age pension
plus	_xx028	Unemployment Benefits from Insurance
plus	_xx029	Sickness Cash benefits
plus	_xx030	Employment Injuries/Occupational diseases benefits
plus	_xx031	Invalidity benefits
plus	_xx032	War related benefits
plus	_xx033	Family Benefits
plus	_xx034	Maternity benefits
plus	_xx035	Government cash transfers for Education

_xx028 Unemployment Benefits from Insurance

include:

- total unemployment
- partial unemployment
- strike benefits

excludes means-tested unemployment benefits

_xx029 Sickness Cash Benefits

includes only sick pay from state insurance
excludes sick pay from employer

_xx030 Employment Injuries/Occupational diseases benefits

Benefits for injured workers
also benefits for surviving spouses and orphans

_xx031 Invalidity Benefits

Benefits from state insurance for former employed individuals with occupational or general invalidity to work

note: in some countries the disability pay is replaced by old age pensions for individuals older than 65

_xx032 War related benefits

Include also benefits for survivors

_xx033 Family Benefits

include:

- child allowances
- benefits for motherless and fatherless child
- education allowances

_xx034 Maternity benefits

include

- prenatal, child birth and post-natal allowances
- benefits for assistance: (baby sitters etc)

_xx035 Government cash transfers for Education

include:

- allocation at reopening of the school year
- scholarships
- state grants for higher education
- retraining grants from labour office

_xx036 TOTAL MEANS-TESTED INCOMES

=	_xx037	Social Assistance
plus	_xx038	Additional Social Assistance
plus	_xx039	Unemployment assistance
plus	_xx040	Other Income Dependant Benefits

_xx037 Social Assistance

fixed monthly payments to guarantee
sufficient income sources; all payments must be means-tested

include also:

- special Assistance for one parent families, if means-tested
- Assistance to cover individual needs in difficult situations through grant of allowances (e.g. illness,care or particular social difficulties)

_xx038 Additional Social Assistance

- means-tested benefits but aperiodic and single payments for heating, clothes,household and urgent needs

e.g. foodstamps, heating allowances
e.g. single grants from welfare associations

_xx039 Unemployment assistance

includes only means-tested unemployment benefits

_xx040 Other Income Dependant Benefits

benefits paid to (low income) households, e.g. housing benefits

_xx041 TOTAL SOCIAL SECURITY INCOME

=	_xx027	Total social insurance income
plus	_xx036	Total means-tested incomes

_xx042 TOTAL PRIVATE INCOME

=	_xx043	Cash Alimony or Child Support
plus	_xx044	Received private Cash Inter-household Transfers

_xx043 Cash Alimony or Child Support

- support from a divorced partner who has to pay for his former spouse and his children
- support from father for illegitimate children
- support from state system for those awaiting alimony or child support from spouses

_xx044 Received private Cash Inter-household Transfers

- from parents in law and children
- one time support from divorced spouses
- from friends and neighbours
- other individuals

_xx045 TOTAL TRANSFERS (EXCL OLD AGE PENSIONS)

= _xx052 TOTAL TRANSFERS
minus _xx022 TOTAL OLD AGE PENSION

_xx046 Transfers for Handicapped

include supplementary allowance for handicapped children and adults (transfers for persons receiving care or care-givers), also benefits for blind civilians

_xx047 Other Transfers

include here transfers which are unqualified or cannot be allocated to previous listed transfer income sources, e.g.

- benefits for parents when they stay at home because of ill children
- marriage grants
- death grants

_xx048 Annuities from insurance

include

- life-annuity from private persons
- regular payments from private insurance

_xx049 Other Lump-sum Payments

report here one time payments such as

- lottery winnings
- inheritances
- big settlement from insurance company

_xx050 Income from home consumption, or from farming and gardening

Note: the income of farmers is reported under _xx011

_xx051 Other Incomes

include here incomes, which are unqualified or cannot allocated to previous listed income sources

_xx052 TOTAL TRANSFERS

=	_xx018	Total employer pension
plus	_xx027	Total social security income
plus	_xx036	Total means-tested incomes
plus	_xx042	Total private income
plus	_xx046	Transfers for Handicapped
plus	_xx047	Other Transfers

_xx053 TOTAL GROSS INCOME

=	_xx017	Total factor income
plus	_xx048	Annuities from insurance
plus	_xx052	Total transfers

_xx054 TOTAL Social security contributions

=	_xx055	Health Insurance
plus	_xx056	Old Age Insurance
plus	_xx057	Unemployment Insurance

_xx055 Health Insurance

includes:

- mandatory employee contribution or
  mandatory contribution for self-employed for:
  state insurance/private insurance

_xx056 Old Age Insurance

includes:

- mandatory employee contribution or
  mandatory contribution for self-employed for: state insurance

_xx057 Unemployment Insurance

mandatory employee contribution

_xx058 Other Direct Taxes

include:

- property and wealth tax
- Church taxes

_xx059 Income tax

personal income tax liabilities

_xx060 TOTAL NET INCOME:

=	_xx053	Total gross income
minus	_xx054	Total social security contributions
minus	_xx059	Income tax

_xx061 Total Contributed Private Cash Inter-household Transfers

=	_xx062	Contributed Transfers to Parents
plus	_xx063	Contributed Transfers to Children
plus	_xx064	Contributed Transfers to Spouses
plus	_xx065	Contributed Transfers to other Relatives
plus	_xx066	Contributed Transfers to others

_xx062 Contributed Transfers to Parents

Transfers to parents and parents-in-law who do not live in the household

_xx063 Contributed Transfers to Children

Transfers to children and stepson/stepchild who do not live in the household

_xx064 Contributed Transfers to Spouses

Transfers to current (separated) spouse who does not live in the household

_xx065 Contributed Transfers to other Relatives

Transfers to other relatives who do not live in the household

_xx066 Contributed Transfers to others

Transfers to other non-relatives who do not live in the household

B.3.4.2. Demographic variables definitions

B.3.4.2.1 Demographic variables definitions on individual level

P201 Sex

Labels
1 male
2 female

P202 Birth year

four digits (e.g. 1940)

P203 Birth month

two digits (e.g. 01 for January)

Pxx204 Family Status

The category married should contain **only those legally married**. Variable Pxx210 (Cohabitor Status) allows us to differentiate between legally married and cohabiting.

Labels
1 married
2 separated
3 single
4 divorced
5 widowed

Pxx205 Relationship to Reference Person in HH

In most cases the reference person (also called "head of household") will be the husband, but since this is not always the case a more general term was chosen.
Adoptive children are treated as legal children and are therefore assigned to category '4'.

note for analysis: also variables PxxL09 (ID-Spouse), PL10 (ID-Father) and PL11 (ID-Mother) can be used to determine in more detail the family relationship between individuals.

Labels
1 Reference Person (RP)
2 Spouse of RP
3 Cohabitor of RP
4 Son, daughter of RP or spouse or cohabitee
5 Foster child of RP or spouse or cohabitee
6 Son/daughter-in-law of RP or spouse or cohabitee
7 Father, mother of RP
8 Father, mother of spouse or cohabitee
9 Brother, sister of RP or spouse or cohabitee
10 Grandchild of RP or spouse or cohabitee
11 Other relatives of RP or spouse or cohabitee
12 Non-relatives of RP or spouse or cohabitee

Pxx206 Foreign Nationals

This variable indicates for every country whether the individual is national or foreign national.

Labels
1 national
2 foreign national

Pxx207 Degree of Handicap

two digits
e.g 50 = 50 % handicapped
80 = 80 % handicapped

00 = no handicap

Pxx208 Visits to Doctor (excluding dentists)

This variable contains the number of visits to doctor by year (excluding visits to dentists).

Pxx209 Subjective Health Status

Following coding scheme has been used.

Labels
1 excellent
2 good
3 fair
4 poor
5 very poor

Pxx210 Cohabitor Status

This variable indicates if two adults living together without being married.

Labels
1 legally married
2 cohabiting
3 other status

B.3.4.2.2 Demographic variables definitions on household level

Hxx250 No. of Adults in HH

Number of persons of 18 or more years.

Hxx251 No. of Children in HH

Number of persons of less than 18 years

Hxx252 Actual age of youngest co-resident child

children are all persons with age less 18

Hxx253 Actual age of oldest co-resident child

children are all persons with age less 18

Hxx254 No. of Adult Children in HH

Number of persons 18 years and less than 30 years who are children of the reference person or of the spouse

Hxx255 No. of Non-resident children

This are children of the reference persons or of the spouse which do not live in the household now.

Hxx256 Household Typology

1 = Single man (1 person household)
2 = Single woman (1 person household)
3 = couple (2 person household)
4 = couple with children
5 = one parent family
6 = three generation household
7 = other households

a) Category '3' and '4':
Couples are legally married couples and cohabitors.

b) Category '4' and '5':
Children are defined here as all persons with age less than 18 years and those persons older than 17 years and less than 30 years who are children of the reference person or of the spouse or the cohabitee (including adoptive children)

c) Category '4':
Couple with children contains only the couple and children and no other persons in the household than children.

d) Category '5':
One parent families are families where the head is living alone with children and where is no cohabitor or the parents of the head and no other persons in the household than children.

d) Category '6':

Three generation household comprises all households in which children, parents and grandparents live together without further persons not belonging to this scheme.
Here children are meant in a wider definition than described for category '4' and '5' above. Children are here all individuals of any age where father/mother is living in the same household. Therefore adults with parents and grand-parents are also included here. The generation, to which the reference person in the household belongs to, does not matter.

Three cases are possible:

(1) Head/Spouse or Cohabitee is living together with**Father/Mother** of Head/Spouse or Cohabitee and with **Grandfather/Grandmother** of Head/Spouse or Cohabitee.

(2) Head/Spouse or Cohabitee is living together with **Children** of Head/Spouse or Cohabitee and with **Father/Mother** of Head/Spouse or Cohabitee.

(3) Head/Spouse or Cohabitee is living together with **Children** of Head/Spouse or Cohabitee and with **Grandchildren** of Head/Spouse or Cohabitee.

As soon as another person (e.g. brother/sisters, uncle/aunts or non-relatives) not belonging to the straightforward definition of three generations:
"children - parents - grandparents" lives in the household, category 7 (other households) has to be chosen.

e) other households are those where non relatives or relatives to whom the previously defined categories do not apply are living in the household.

B.3.4.3. Labour Force and Work history variables definitions

Pxx301 Employment Status

Labels
1 under 16 years
2 student
3 working
4 unemployed
5 housewife
6 retired
7 other

Pxx302 Employment Type

- overtime excluded -

1 full time (30 hours or more)
2 part time (20 - 29 hours)
3 some hours (less than 20 hours)

Labels
1 full time
2 part time
3 some hours

Pxx303 Prof.Stat.(long version)

This variable is currently available only for Germany and Luxembourg. For the short version (collapsed values) including the USA, see Pxx304.

Labels
11 farmers
12 self employed
13 contributing family workers
21 blue collar
22 white collar
23 apprentice
31 civil servants

Pxx304 Prof. Stat.(short version)

The data for Germany and Luxembourg are derived from Pxx303 by

COMPUTE Pxx304=TRUNC(Pxx303/10).

1 self employed
includes:

- self employed
- farmers
- contributing family workers

2 employees
includes:

- blue collar
- white collar
- apprentice

3 state sector
includes:

- state sector (USA)
- civil servant (Germany, Lux)

Labels
1 self employed
2 employee
3 state sector

Pxx305 Normal Working Hours Weekly

Agreed upon amount of weekly working time (hours)

Pxx306 Total Working Hours

Actual time worked per week, including possible overtime (on average, hours per week)

Pxx307 Overtime Compensation

Compensation by payment or free time
(not available in PSID)

1 payment
2 time off in lieu
3 payment and time off in lieu
4 no compensation
5 no payment

Pxx308 Second Job

Whether person has second job or not

Labels
1 yes
2 no

Pxx309 Working Hours Second Job

Working hours per week on second job (on average)

Pxx310 Employment Sector

Use 1 digit ISIC coding scheme from year 1988.

1 = agriculture, hunting, forestry and fishing
2 = mining and quarrying
3 = manufacturing
4 = electricity, gas and water
5 = construction
6 = wholesale and retail trade and restaurants and hotels
7 = transport, storage and communication
8 = financing, insurance, real estate and business service
9 = community, social and personal services
0 = activities not adequately defined

Pxx311 Type of Employer

1 = public employer
2 = no public employer

Private employers are defined here as all those companies which have the primary target to make profits.
Public employers are all those establishments which are run by the state and which have the primary target to produce services and do not make profits necessarily.

'Nationalized industries' or other companies which are controlled by the state are defined here as private employer.Private non-profit orgainzations has been counted as public employer.

Pxx312 Firm Size

convert brackets of classes of employees using midpoints into quantitative values

Pxx313 Type of Current Occupation

Use ISCO 2 digits coding scheme from year 1988.

Pxx314 Contractual Situation

1 = contract - Permanent
2 = contract - Fixed Term
3 = No contract

Pxx315 Total Month in Employment in Most Recent Job

Pxx316 Employment Status by Year (Retrospective)

1 = student
2 = full time working
3 = part time working
4 = unemployed
5 = housewife
6 = retired
7 = other

This variable will be created for Germany and UK by using
biographical information collected with retrospective questions. For the othercountries this

variable will be created by using all available waves with data.

Pxx317 Labour Force Status"

Labels
1 working now
2 not working

This variable reflects the precisely defined labour force status at one specific reference date. The status "Working" comprises all individuals that are working at least one hour per week (apprentices included). The relevant working hours for this classification are the sum of working hours for first and second jobs. Individuals which classify themselves in variable Pxx301 as "Unemployed","Retired", "Other" but have small, casual or second jobs, have been recoded into the category "Working" in Pxx317. An indication if somebody has a small job etc. might be also the existence of earnings at the reference date.

Individuals that have a job, but are temporarily away from their job because of maternity leave, holiday, sickness, training courses etc. are also classified here as working.

Pxx318 Unemployment Status

This variable gives information about the fact of being unemployed and registered at the labour office or not.

Labels
1 registered
2 not registered

This variable has only valid information for those individuals which have the status "unemployed" in variable Pxx301.

Pxx319 Immediate Intention for Work

Labels
1 yes
2 no

It is assumed that those individuals who want to re-enter immediately the labour market gave the answer " yes" to following questions:

e.g.: Have you looked for any kind of paid work in the last week or in the last four weeks ?
e.g.: Have you been doing anything in the last four weeks to find a job?
e.g.: Do you want to enter on work immediately ?

This variable has only been created for individuals with status "non working" in Pxx317

and which have the category "no" in variable Pxx320.

Pxx320 Future Intention for Work

Labels
1 yes
2 no

It is assumed that those individuals want to re-enter the labour market in future (not immediately) which gave the answer " yes" to following question:

e.g.: Are you thinking of getting a job in the future", under the condition that they have not been active in the last four weeks to find a job

e.g.: Do you want to enter on work not immediately, but later in the next years ?

This variable will be only created for individuals with status "non working" in Pxx317 and which have the category "no" in variable Pxx319.

B.3.4.4. Education variables definitions

The highest education level that a person ever obtained during the waves which are included in the PACO dataset is available in the Person Inventory files and in each yearly related individual file(P401,P402,P403). The current education level is included in the year-related files only (Pxx401, Pxx402, Pxx403).

P401 Highest obtained School Education

According to OECD-Classification

1 First level (primary):

For all countries 1st to 6th grade (Germany: to 4th grade)

2 Second level --- first stage

- corresponds in many countries to end of obligatory education
- in most cases general education, including in France, Greece, the Netherlands a so-called pre-vocational training, not really related to a specific occupation

3 Second level --- second stage

includes:

- general education
  preparation for university or other third level education not directly leading to a profession
- technical / occupational / vocational education leading to occupation or group of occupations
- apprenticeship

4 Third level (tertiary education)

includes:

- university
- technical college or institute

Labels
1 First level
2 Sec. lev., first stage
3 Sec. lev., sec. stage
4 Third level

P402 2nd Lev. 2nd Stage Education

In addition to variable P401 the second level second stage education is split in order to distinguish between academic and professional or technical education.

1 general education
preparation for university or other third level education not directly leading to a profession

2 technical / professional education leading to profession or group of professions

3 apprenticeship

Labels

1 high school
2 professional education
3 apprenticeship

P403 Imputed Years of Education

Number of years of education

- includes first to third level education (see P401)

Technical remark:

In most cases the number of years is not available from the datasets but only the type of school degree. In these cases the minimum number of years necessary to achieve this degree is assumed, except for the lowest level (no degree), where it is assumed that only one year was missing.

Pxx401 Current School Education
Pxx402 Current 2.lev 2.stage Education
Pxx403 Current years of education
see definitions of variables P401, P402, and P403 above

B.3.4.5. Housing variables definitions

Hxx601 House ownership Status

1 Owner
2 Tenant
3 Living Rent Free

B.3.4.6. Other variables definitions

B.3.4.6.1 Other variables definitions on household level

Hxx801 Territorial division

Use NUTS CODE of the EEC (Nomenclature of territorial units for statistics).

Give information on lowest possible level.

Character variable with length 5

Hxx813 Child care external to household

time unit: hours per week

B.3.4.6.2 Other variables definitions on individual level

Pxx810 Time spent on unpaid housekeeping work

e.g. cooking and cleaning

time unit: hours per week

Pxx811 Time spent on unpaid child care

time unit: hours per week

Pxx812 Time spent on other unpaid care

including volunteer work

time unit: hours per week

B.3.4.7. Weighting variables definitions

PxxWEIG Individual Weight
cross-sectional weight
see chapter B.4.3

PxxWEIX Individual Weight
Cross-sectional weight for Luxembourg, to be used when the extension of the sample, which was added in 1991 is analyzed.
see chapter B.4.3

PxxWEIL Individual Longitudianl Weight
longitudinal weight from wave 1 to wave t (year xx)
see chapter B.4.3

HxxWEIG Household Weight
cross-sectional weight
see chapter B.4.3

HxxWEIX Household Weight
Cross-sectional weight for Luxembourg, to be used when the extension of the sample, which was added in 1991 is analyzed.
see chapter B.4.3

PxxPROB Individual Retention Probability

The retention probability is the conditional probability of remaining in the panel after each selection step, i.e.after each wave.
see chapter B.4.3

B.3.4.8. Organizational (Link) variables definitions

_L01 Country

Character variable with length 2, contains:

'fr'	France
'ge'	Germany
'hu'	Hungary
'lu'	Luxembourg
'pl'	Poland
'uk'	UK
'us'	USA

_L02 Year

four digits (e.g. 1985)

_xxL03 ID-Household

unique household identifier, should remain constant between years, if no split-off occurs.

If the original panel study does not use constant household identifiers new articifical household identifiers have been created. If splits occurs it must be defined which is the main household (keeping the old identifier) and which is the split off household (getting an new identifier). The household in which either the reference person(1.priority) or the spouse (2.priority) lives is to be defined as main household. If a couple divorces or separates, the household of the husband will be the main household and the household of the divorced wife will be the split off.

_xxL04 Pre-Year ID-HH

contains the household identifier of the previous year, in most cases _xxL04 = _xxL03. Only in the case of a split-off _xxL03 is different from _xxL04.

_xxL05/_L05 ID-Group

only Lorraine/Luxembourg/USA: contains the identifier of the income groups. The identifier must be unique within one household.

identifier 1 = first income group
2 = second income group
3 = third income group

For other country datasets two artificial (income) groups can be build. The first group comprises the reference person and the spouse (cohabitant) of the reference person. The second group would contain the income of all other individuals in the household.

PL06 ID-Individual

unique individual identifier between all individuals, must remain constant between all years.

_xxL07 ID-Reference Person

This identifier is a pointer to the reference person in the household. In most cases this is a pointer to the head of the household. Because the reference person can change between waves, therefore PxxL07 may also change

between years.

_xxL08 ID-Spouse of Reference Person

This identifier is a pointer to the spouse of the reference person. Spouses are defined here as **legal spouses and cohabitors.** PxxL08 may change between years.

PxxL09 ID-Spouse

This identifier is a pointer to the spouse. Spouses are defined here as **legal spouses and cohabitors**. PxxL09 may change between years.

PL10 ID-Father

This identifier is a pointer to the father who lives now or lived before (within panel years) also in the household. PL10 is **not** allowed to change between years.

PL11 ID-Mother

This identifier is a pointer to the mother who lives now or lived before (within panel years) also in the household. PL10 is **not** allowed to change between years.

_L12 Case ID

The case-ID is the household number of the first wave of a panel study. If split-offs appear, the original household and the split-off household still will have the same case-id, but the current houshold number _xxL03 will be different for original and split-off households.

_L13 Random Group ID

By the Random Group ID the whole sample is split into subgroups.The wave 1 household sample is split at random into groups of the same size. Split-off households and new panel members belong to the subgroup of the household they stem from. Thus each subgroup can be regarded as a subpanel.

The Random Group ID is useful for variance estimation and hence for the estimation of confidence intervals, where other methods for variance estimation are not easily available.

The random group ID can be used for the random groups method or for the jackknife technique.

Appendix A explains how this ID variable is created.

_xxL14 Match Indicator

0 incomplete information (unit nonresponse)
1 complete information

This variable is only necessary for the datsets for Germany and the US. For these datasets information from following waves (t+1),(t+2) must be matched with information form wave(t) in order to create the correct dataset for year (t). Variable _xxL14 will be set to '1' if we have information from all necessary waves (complete information). A value of '0' will be set if at least one information from a following wave is missing (unit response).

The availibility of this match variable together with the information of missing values will allows us to differenciate between item non response and unit nonresponse.

B.4 Description of Methodological Variables

B.4.2 Use of Random Group Indicators

The Random Group ID is useful for variance estimation and hence for the estimation of confidence intervals, where other methods are not easily available.

The variables PL13 and HL13 are supplied on the individual and household level, respectively. These variables remain unchanged, thus there is no wave indicator in the variable name.

By the Random Group ID the whole sample is split into subgroups. This is done by splitting the wave one household sample at random into groups of the same size (the method is described in detail below). Split-off households and new panel members belong to the subgroup of the household they stem from or into which they enter. Thus each subgroup can be regarded as a subpanel.
The idea is to estimate a parameter of interest within each group separately and to estimate the variance between these group estimates.
The number of random groups R should not be to small, in order to estimate the variance on the basis of a large number of observations. On the other hand the sub-groups should not be to small.
The number R = 8 was chosen in the German Socio-economic panel because of design reasons and turned out to be a practical size.
The same number was chosen for the other studies included in PACO.

B.4.2.1 The theory

The underlying theory proceeds on the assumption that we have several (say R) independent samples of equal size, drawn with the same sampling design. From each sample a separate estimate of the parameter of interest Y is calculated. The sample variance is calculated among these estimates (for details see Wolter, 1985).

Definitions

Y	population parameter
$\hat{Y}_1,...,\hat{Y}_R$	estimates obtained from the random groups with $E(\hat{Y}_r) = \mu$
$\overline{\hat{Y}} = \frac{1}{R} \sum_{r=1}^{R} \hat{Y}$	overall estimation of Y

There are two alternatives to estimate the variance of $\overline{\hat{Y}}$:

$$V_1 \;=\; V_1(\bar{\hat{Y}}) \;:=\; \hat{s}_R^2 \;=\; \frac{1}{R\,(R-1)} \sum_{r=1}^{R} (\hat{Y}_r - \bar{\hat{Y}})^2$$

$$V_2 \;=\; V_2(\bar{\hat{Y}}) \;:=\; \hat{s}_R^2 \;=\; \frac{1}{R\,(R-1)} \sum_{r=1}^{R} (\hat{Y}_r - \hat{Y})^2$$

The difference is constituted by the last term. In V_1 the reference is the mean of the estimators obtained from the random groups, whereas in V_2 the whole sample is used for the estimation.

Under the conditions described above (independent samples) the estimations have the following properties:

$E(\hat{Y}_r) = \mu$ (not necessarily Y)
$E(V_1) = Var(\bar{\hat{Y}})$, because $\hat{Y}_1,...,\hat{Y}_R$ are uncorrelated
$V_1(\bar{\hat{Y}}) \neq V_2(\bar{\hat{Y}})$, in general
$V_1(\bar{\hat{Y}}) = V_2(\bar{\hat{Y}})$, for linear estimators

B.4.2.2 The application to PACO

<u>The random group estimates</u>

In the case of an existing panel which is split into subgroups, the necessary conditions - as were described above - hold only approximately because the independence property is violated.
In the case of equal sized groups we have

$$E(V_1) = Var(\bar{\hat{Y}}) - Cov(\hat{Y}_i,\hat{Y}_j) \quad .$$

But it was shown by Wolter (1985) that the bias by the covariance will be relatively small in large surveys and small sampling fractions, which is the case in the panel studies included in PACO.

Hence the variance of a population estimate can be estimated by

$$V_1 = V_1(\bar{\hat{Y}}) := \hat{s}_R^2 = \frac{1}{R\,(R-1)} \sum_{r=1}^{R} (\hat{Y}_r - \bar{\hat{Y}})^2 \qquad \textbf{(1)}$$

and a confidence interval can be set up in the usual way:

$$CI_{\hat{s}_R} := [\bar{\hat{Y}} \pm \hat{s}_R \cdot t_{R-1,1-a/2}] \qquad \textbf{(2)}$$

Another way of estimating a confidence interval is based on the ordered values

$\hat{Y}_{(1)},...,\hat{Y}_{(R)}$, where $\hat{Y}_{(1)}$ is the smallest and $\hat{Y}_{(R)}$ is the biggest value.

If the median of the distribution of $\hat{Y}_r$ is equal to Y, which means that the distribution is symmetric around Y, the following intervals can be created for R = 8, which is the number chosen for PACO (see Büning, Trenkler, 1978):

$$\begin{aligned} P(\hat{Y}_{(1)} \le Y \le \hat{Y}_{(8)}) &= 1 - 0.008 \\ P(\hat{Y}_{(2)} \le Y \le \hat{Y}_{(7)}) &= 1 - 0.07 \\ P(\hat{Y}_{(3)} \le Y \le \hat{Y}_{(6)}) &= 1 - 0.29 \end{aligned} \qquad (3)$$

The second equation leads to a 93% confidence interval $[\hat{Y}_{(2)},...,\hat{Y}_{(7)}]$ which is the closest one to the usual 95% interval.

The Jackknife Method

The random groups can be used for Jackknife estimation as well.

For a Jackknife estimate of a parameter Y we calculate the parameter estimates

$\hat{Y}_{(-r)}$, $r = 1,...,R$, which are computed by using all data except for the r^{th} group.

From these values the so-called pseudo values are calculated :

$$\hat{Y}_r := R\hat{Y} - (R-1)\,\hat{Y}_{(-r)} \quad ,$$

and the Jackknife estimate of Y is the mean value of these pseudo values

$$\bar{\hat{Y}} = \frac{1}{R} \sum_{r=1}^{R} \hat{Y}_r \quad .$$

This estimate has the property that it reduces bias.
The variance can be estimated in an analogous way to the random group version, using the pseudo values instead of the estimates obtained directly from the random groups.

B.4.2.3 Difficulties in the achievement

As it was mentioned above, each study included in PACO was split into 8 random groups. Yet the procedure of splitting the whole sample on the basis of the households of the first wave could not be applied strictly in any case.

PSELL - Luxembourg

In Luxembourg an extension of the sample was added in wave 7 (1991). The whole sample was drawn in 1985 but only part of the households were interviewed in 1985, the rest was added in 1991. The basis for the splitting into random groups is the whole sample of 1985 with the extension in 1991. A technical difficulty which had to be over-come, was the fact that there were split-off households in the extension in 1991, which had to be put into the same random group as their household of origin. There were also split-off households of which the original household did not participate in 1991.

PSID - USA

The PSID data are included from 1983 onwards. For reasons of practicability the splitting into random groups was performed on the basis of 1983. Hence it is not guaranteed that households that split from their original households before 1983
are kept in the same random group as their original household. After 1983 the rule is fulfilled.

ESML - France / Lorraine

In France the wave one sample was taken from the pre-test and in wave two a large amount of households was added. Thus the combination of these waves is the basis for the splitting mechanism.

BHPS - UK

In the BHPS a few households that could not be interviewed in wave one were interviewed in wave two for the first time. They were added to the basis on which the splitting was performed.

B.4.2.4 The Split Algorithm

First a file of all original households excluding split-off households is set up. Normally this should be the wave one household file (see exceptions above). The original household ID is stored in the variable HL12.
The file can be split at random into 8 subgroups by using the following SPSS statements:

```
compute help=uniform(1000).
sort cases by help.
compute hl13=mod($casenum,8)+1.
sort cases by hl12.
save outfile=rgroup/keep=hl12 hl13.
```

This file can be matched to household and individual files by HL12.

B.4.2.5 Examples of Application

For the measures of concentration like the Gini- or Theil- coefficient a formula for the estimation of a confidence interval is not easily available. By the random groups method confidence intervals can be computed easily.
The following program-statements calculate a Gini-coefficient for each year for Luxembourg (1985 - 1989) and the USA (1983 - 1987). For each year the results are written to a file which is displayed afterwards in the order of the size of the estimates. The second and seventh value can easily by taken to set up the confidence interval.

```
define !ginimac
       (start=!token(1)
        /end=!token(1)
        /country=!token(1)
        /rg1=!token(1)
        /rg8=!token(1)).
* loop years.
!do !xx=!start !to !end
!let !hxx017=!concat('h',!xx,'017').
!let !hxxweig=!concat('h',!xx,'weig').
*.
* Files for Luxembourg.
!if (!country=LU) !then
!let !pluxxh=!concat('c:\paco\lux\plu',!xx,'h.sav').
!let !pluh=!concat('plu',!xx,'h').
file handle !pluh /name=!quote(!pluxxh).
!let !gluxx=!concat('c:\paco\lux\glu',!xx,'.sav').
!let !glu=!concat('glu',!xx).
file handle !glu /name=!quote(!gluxx).
!ifend
*.
* Files for the USA.
!if (!country=US) !then
!let !pusxxh=!concat('c:\paco\usa\pus',!xx,'h.sys').
```

```
!let !push=!concat('pus',!xx,'h').
file handle !push /name=!quote(!pusxxh).
!let !gusxx=!concat('c:\paco\usa\gus',!xx,'.sav').
!let !gus=!concat('gus',!xx).
file handle !gus /name=!quote(!gusxx).
!ifend
*.
*.
* loop random groups.
!do !rg=!rg1 !to !rg8
!let !ginirg=!concat('gini',!rg).
!if (!country=LU) !then
get file =!pluh /keep=hl13 !hxx017 !hxxweig.
!ifend
!if (!country=US) !then
get file =!push /keep=hl13 !hxx017 !hxxweig.
!ifend
compute rg=number(!quote(!rg),f1.0).
*.
* select random group rg .
select if hl13=rg.
* calculate the gini-coefficient within random group rg.
compute y = !hxx017.
compute hweight = !hxxweig.
recode y (low thru 0=.1).
sort cases by y (a).
compute cumwgt=cumwgt + hweight.
leave cumwgt.
compute temparea = y * cumwgt.
compute group= 1.
weight by hweight.
aggregate outfile=*
   /presorted
   /break=group
   /rgr=max(rg)
   /n=n
   /sumy sumarea = sum(y temparea).
compute gini = (2*sumarea/sumy - n - 1)/n.
print formats gini (f10.4).
* list variables = gini.
save outfile=!ginirg.
!doend
* loop random groups.
add files file=gini1
         /file=gini2
         /file=gini3
         /file=gini4
         /file=gini5
         /file=gini6
         /file=gini7
         /file=gini8.
```

```
compute gini=gini*100.
* sort cases by gini.
execute.
list variables=gini.
descriptives variables=gini.
*.
!doend
* loop years.
!enddefine
!ginimac start=85 end=89 country=LU rg1=1 rg8=8.
!ginimac start=83 end=87 country=US rg1=1 rg8=8.
```

The following table lists the gini-coefficients calculated within the random groups for Luxembourg in 1985 and 1989 and their differences within the random groups. The differences are sorted, and the 93 % - confidence interval (formula 3) results in (-0.19 % ,3.79 %).

Year	1985	1989	Difference
Random Group			
1	51.75	51.93	-0.19 (2)
2	49.48	46.66	2.82 (6)
3	46.86	48.19	-1.33 (1)
4	49.87	43.22	6.64 (8)
5	45.69	44.34	1.35 (4)
6	42.98	41.24	1.74 (5)
7	51.24	49.96	1.28 (3)
8	50.41	46.63	3.79 (7)
entire sample	46.52	48.53	2.01

B.4.3 Weighting Variables

The following paragraph gives an overview on the weighting variables that are available in the PACO dataset.
Although it is the aim of the PACO project to harmonize and standardize the variables of the domestic panels, the weight variables cannot be harmonized a posteriori, since different sampling designs and weighting methods have been used in the national studies.
In paragraph 4.3.2 the methods which have been employed in the domestic panel studies and their main differences are briefly described.
Detailed descriptions of the surveys and weighting methods can be obtained from the manuals of the national panel studies and / or from additional literature.

B.4.3.1 Overview of Available Weights

The following table gives an overview of the weights which are included in the panel datasets and which have been included in the PACO dataset.

Variable name	Dataset	original Panel	Remarks
replace _ by P for individuals, H for households			
cross-sectional weights			
_xxWEIG	PGExx_ PLUxx_ PUKxx_	SOEP PSELL BHPS	**PSELL:** appropriate until 1990 (see _xxWEIX)
	PUSxx_	PSID	**PSID, individuals**: PUSxxP applies only to sample members from wave one, which means that these weights are in fact longitudinal ones; if all persons are to be analyzed, take the household weight, which is more appropriate except for the rising factor (see *Rescaling* below).

Variable name	Dataset	original Panel	Remarks
_xxWEIX	PLUxx_	PSELL	In 1991 an extension was added to the sample. _xxWEIX should be used, when the sample including the extension is analyzed.
longitudinal individual weight			
PxxWEIL	PLUxxP	PSELL BHPS	longitudinal individual weight wave 1 --> t (year xx)
individual retention probability for the construction of longitudinal weights wave t-1 --> t (t = xx)			
PxxPROB	PGExxP	SOEP	These variables can be used to construct longitudinal weights for periods starting in any wave (thus new entrants can be included in the analysis). In order to calculate a longitudinal weight for the period from wave s to wave t, the cross-sectional weight for wave s can be multiplied by the reciprocals of the appropriate retention probabilities. Example : longitudinal weight for period 1986 -1990: P86WEIG*P87PROB*P88PROB*P89PROB This variable is not available in studies, in which the response rates are not calculated on the base of wave t -1 but on the base of the starting year. In the BHPS it is not supplied.

Rescaling

The weights in the PSELL and BHPS are rescaled to sum up to the sample size.
The weights in the SOEP are rescaled to sum up to the population size.
The weights in the PSID are rescaled to sum up to one percent of the population size.

These differences do not affect percentages, but they have to be taken into account whenever standard deviations and especially standard deviations of the mean are calculated by standard software. SPSS and SAS treat the weights differently.
Anyway the calculation of the variance or standard deviation in a panel is not a straightfor-

ward procedure,which is obvious in the case of longitudinal analyses.
In the case of cross-sectional analysis it is due to dependent observations within the households and maybe even between households.
In paragraph B.4.3 an alternative variance estimation procedure is described.

B.4.3.2 Overview on the weighting methods employed by the panel studies

The weighting procedures of the first wave differ because of the different sampling methods. In any case the weights adjust for the different selection probabilities of the households.
But also the adjustments for attrition from wave two onwards have been made in different ways.

2 One of the differences is given by the different treatment of so-called **non-sample persons**, which influences not only the following rules but also the weighting of households with new entrants.

3 The **level of adjustments** for non-response differ and

4 so does the **wave** which is taken as the base for the adjustments.

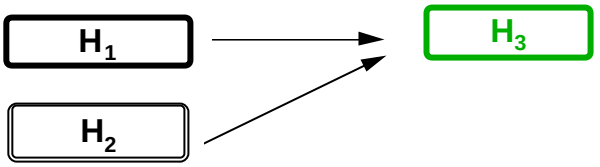
5 Furthermore the **statistical methods** to estimate response probabilities differ. Logit models are used in Germany from wave 2 on , whereas in the USA, in Luxembourg, and in the British panel, a partition of the sample is chosen by the A.I.D. method.

6 The **treatment of households with new entrants**, so-called non-sample persons is different.

1) Following Rules	
PSELL, PSID, BHPS Only sample-persons, which means panel members of the first wave and their descendents are followed, if they leave a panel household. If all sample persons have left the household the remaining persons are not interviewed.	**SOEP** No distinction is made between sample persons and none sample persons. Once a person has entered a panel household, he or she is treated as panel member. This rule was applied from wave five on-wards.

Weighting of the Waves t > 1	
2) Level of Adjustment	
PSELL, BHPS, PSID : Individuals **<u>Calculation of household weight</u>** **PSID :** average of the weights of the head of the household and his / her spouse **(PSELL, BHPS):** average of the weights of the adult sample persons	**SOEP: Households** **<u>Calculation of individual weight</u>:** house-hold weight

3) Reference Wave	
Wave 1	**Wave t - 1**
PSELL, PSID	**SOEP, BHPS**
4) Statistical Methods	
A.I.D. **automatic interactive detection**	**Logit Model**
PSELL, PSID, BHPS	**SOEP**
partition of the sample into subgroups in order to find a set of variables and their interactions which best explain differences in response rates	The binary variable response / no response is explained by characteristics of the units by means of a model.
estimated response probability: response rate within a subgroup	**estimated response probability:** estimated by the model

5) Treatment of households with new entrants	
PSELL , BHPS : Weight share approach step 1 estimation of the individual weights of the sample persons w_i , i=1,...,s (**s :** number of sample persons) step 2 calculation of the initial household weight by the average of the individual weights of the sample persons $w_h^{ini} = 1/s \sum_{i=1}^{s} w_i$ step 3 calculation of the final household weight by reducing the initial weight by a factor : (**n** number of non-sample persons) $w_h^{final} = \frac{s}{s-n} w_h^{ini}$ step 4 assignment of the household weight to the individuals within the household	**SOEP : Estimation of the inclusion probability of the resulting household** **Idea :** The household H_3 had 2 chances to belong to the sample: either by H_1 or by H_2 as panel household in wave t-1. The weight is reduced appropriately:  H1 : panel household in wave t-1 H2 : household which the new member stems from H3 : resulting household P1: estim. inclusion probability of H1 (reciprocal of the weight) P2 : estim. inclusion probability of H2 (estimated by means of individual characteristics of the new member) P3 : P1 + P2 **w3 := 1 / P3 = 1 / (P1 + P2)** step 1 adjust the weight of the household without the new entrant H1 (reciprocal P1) step 2 estimate the inclusion probability P2 step3 combine both to w3

B.5 Available PACO Variables

a) The File "VARLIST" contains the availability of variables by country, year and level.

b) Following files contains the availability of variables by year and level separate by country:

"FRANCE"
"GERMANY"
"HUNGARY"
"LUX"
"Poland"
„Spain"
"UK"
"USA"

c) Explanation of Codes (for files above)

H: Household level
G: Group level
I: Individual level
R: Reference Person
Sp: Spouse
O: Other individual

DCC Definition Comparability Code

1: o.k.
2: minor problems
3: major problems
4: not yet comparable
5: not comparable

LCC Longitudinal Comparability Code

1: o.k.
2: minor problems
3: major problems
4: not yet comparable
5: not comparable

d) The codebook for the different PACO files can be found in following Word files:

household files: H.DOC
group files: G.DOC
individual files: P.DOC

person inventory file: PI.DOC

person work biography file: PB.DOC
person family biography file: PF.DOC

B.6 Overview on PACO Files

B.6.1 Structure of PACO DATA BASE

The PACO data are held in a relational data structure (see figure 1). The PACO DATA BASE contains two sets of files: the PACO (result) files and the shadow files. The PACO (result) files hold the standardized PACO variables under identical variable names and data structures for each country. The shadow files contain a subset of not standardized variables from the original panel studies under original variable names.

Figure 1: Structure of the PACO DATA BASE for one domestic Panel

	Person inventory file	Cross-sectional Files reference year				Longitudinal Biographical Person file
		85	86	87	xxh	
PACO Files:						
Household		85h	86h	87h	xxh	
Group		85g	86g	87g	xxg	
Person	pi	85p	86p	87p	xxp	pb
Shadow Files:						
Household		85h	86h	87h	xxh	
Group		85g	86g	87g	xxg	
Person		85p	86p	87p	xxp	

Each separate PACO (result) file with standardized variables contains data for one specific country. The cross-sectional files hold data for one reference year and may contain data collected in different waves.

The cross-sectional files are available for following units:

- Household (xxh)
- Income Group (xxg)
- Person (xxp)

The cross-sectional file for persons (xxp) contain records for all individuals (adults and children) in the household. The majority of variables for children records will be missing.

In addition to the cross-sectional files following two other files are available:

- Person Inventory file (pi)
- Longitudinal Biographical File (pb)

The (static) Person Inventory file (pi) contains constant variables for individuals as sex, birth year and the identifiers of the father and mother. The (longitudinal) Biographical File (pb) contains the biographical vectors with retrospective information about education and labour force status for individuals.

B.6.2 Convention of PACO file names

The names of each file in the PACO Data Base are defined according following scheme.

Year related files: fccxxu
Non-year related files: fccuu

The first character (**f**) of each file name specifies the file type, two characters (**cc**) the country, two characters (**xx**) the reference year and the last character(s) (**uu**) the target files. Following abbreviations are used:

f denotes the file typ

p = **p**aco result file s = Paco **s**hadow file

cc denotes the country

fr = **fr**ance ge = **ge**rmany hu=**hu**gary pl= **p**o**l**and lu= **lu**xembourg
uk = **u**nited **k**ingdom us = **us**a

xx denotes the year

85 = 19**85** 86 = 19**86** ... **xx** = 19**xx**

u or **uu** denotes the target files for PACO

h = **h**ousehold g = income **g**roup p = **p**erson pi = **p**erson **i**nventory
pb = **p**erson **b**iography

Examples for file names:

plu85h: **p**aco result file for **lu**xemburg in year 19**85** on **h**ousehold level

pfr86p: **p**aco result file for **fr**ance in year 19**86** on **p**erson level

pukpi : **p**aco result file for **uk** holding the **p**erson **i**nventory file

B.6.3 Data structure of PACO (result) files

The standardized variables in all the PACO files are classified here into two categories of variables:.

- analysis and weighting variables
- organizational variables

All PACO result files contain the same number of variables. When it was not possible to create a specific PACO variable from one country dataset 'dummy' variables were created instead. The constant value '-3' (variable is missing) was assigned to these 'dummy' variables.

The total number of maximum available variables is shown in table 1.

Table 1: Number of variables in the PACO Data Base files					
	Person inventory file	Cross-sectional Files			Longitudinal Biographical Person file
		Household files	Group files	Person files	
File name	PccPI	PccxxH	PccxxG	PccxxP	PccPB
organizational variables	4	9	7	14	2
analysis and weighting variables	6	144	132	170	93
Total number of variables	10	153	139	188	95

Table 2: List of analysis variables in the PACO Data Base files					
	Person inventory file	Cross-sectional Files			Longitudinal Biographical Person file
		Household files	Group files	Person files	
File name Domaine	PccPI	PccxxH	PccxxG	PccxxP	PccPB
Income variables _xx001 - _x066 _xxi001 - _xxi066		x	x	x	
Demographic variables p201-p203	x			x	
Demographic variables pxx204 - pxx209				x	
Demographic variables Hxx250 - Hxx256		x			
Labour Force variables pxx301 - pxx320				x	
Employment Status by year pxx316					x
Highest obtained Education p401 - p403	x			x	
Current Education pxx401 - pxx403				x	
Housing: Hxx601		x			
Territorial sub- div: Hxx 801		x			
Time use: pxx810 - pxx813				x	
Child care: Hxx 813		x			
Weights: _xxweig _xxprob		x x		x x	

Table 3: List of organizational variables in the PACO Data Base files					
	Person inventory file	Cross-sectional Files			Longitudinal Biographical Person file
		Household files	Group files	Person files	
File name	PccPI	PccxxH	PccxxG	PccxxP	PccPB
a) Key variables					
_l01 (country)	x	x	x	x	x
_l02 (year)		x	x	x	
_xxl03 (id-household)		x	x	x	
_xxl04 (pre-year id-household)		x	x	x	
_xxl05 (id-group)			x	x	
_pl06 (id-Person)	x			x	x
_xxl07 (id-reference person)		x		x	
_xxl08 (id-spouse of ref. person)		x		x	
pxxl09 (id-spouse)				x	
pl10 (id-father)	x			x	
pl11 (id-mother)	x			x	
_l12 (case-id)		x	x	x	
b) Secector variables					
_l13 (Random Group-id)		x	x	x	
_xxl14 (Match indicator)		x		x	

The table 2 lists the analysis and weighting variables in more detail and shows in which PACO files they can be found.

The table 3 lists the organizational variables and the files where they can be found. In this table the organizational variables are further subdivided into the categories **key variables** and **selector variables**. Keys are variables or a set of (link) variables that identify the cases (records) in a specific file. The key variables are necessary to match and to aggregate the PACO files. Another set of variables is classified here as selector variables. Selector variables are variables used to identify specific cases in a file, but are not really used to match or to aggregate files.

The keys can be classified as **primary** and as **secondary** keys: **Primary keys** are variables that uniquely identify all cases of a specific file. **Secondary keys** are variables that identify a subset of cases in a specific file or specify uniquely cases in another file.

All files of the PACO DATA BASE are pre-sorted by the primary key. The table 4 lists the defined primary key defined for each PACO file.

Table 4: Primary keys and sort order of the PACO Data Base files

Person inventory file	Cross-sectional Files			Longitudinal Biographical Person file
	Household files	Group files	Person files	
PccPI	PccxxH	PccxxG	PccxxP	PccPB
Primary key and sort order				
pl06	hxxl03	Gxxl03/Gxxl05	pl06	pl06

Table 5: Secondary keys in the PACO Data Base files

Person inventory file	Cross-sectional Files			Longitudinal Biographical Person file
	Household files	Group files	Person files	
PccPI	PccxxH	PccxxG	PccxxP	PccPB
Secondary keys				
Pl10 Pl11	Hxxl04 Hxxl07 Hxxl08 Hl12	Gxxl04 Gl12	Pxxl03 Pxxl04 Pxxl07 Pxxl08 Pxxl09	

Table 6: Selector variables in the PACO Data Base files

Person inventory file	Cross-sectional Files			Longitudinal Biographical Person file
	Household files	Group files	Person files	
PccPI	PccxxH	PccxxG	PccxxP	PccPB
Selector variables				
-	Hl13 Hxxl14	Gl13	Pl13 Pxxl14	-

Figure 2 shows the relational structure of the PACO Database indicating the cross-sectional relations. Each box in this figure represents one Database file. The lines between the boxes illustrate the relations (links) between the files. The symbols "n" and "1" attached to the lines define the type of relation. The variable names in the boxes specify the key variables to be used to define the relations. Primary keys are those which are attached to a "1" of the relation lines, secondary keys are those which are attached to a "n".

a) The relation (1:n) between the household file (xxH) and the person file (xxP) express that one household may have one or more household members (n) and that each person is only a member in one household (1). For matching these two files the primary key Hxxl03 (Household identifier) from the household file and the secondary key Pxxl03 (household identifier) from the person file must be used.

b) The relation (1:1) between the person file (xxP) and the person inventory file (PI) tells that each person has only one entry with information in the person file (1) and only one other entry with information in the person inventory file (1). Both files are linked together by the unique person identifier (PI).

The following figures (3-9) show additional relations between the PACO files. The additional possible cross-sectional links from household file to the reference person and to the spouse of reference person are explained in figure 3; the provided links within the person files to other family members can be found from figure 4. Figure 5 displays the longitudinal links between the person files. The three possibilities to match in a longitudinal manner the household files can be found from figure 6-8. The last figure 9 shows the links for a longitudinal match of person files combined with a cross-sectional match between household and persons files.

Figure 2: Relational structure of the PACO Database indicating the cross-setional relations between the different files and the key variables to combine the files

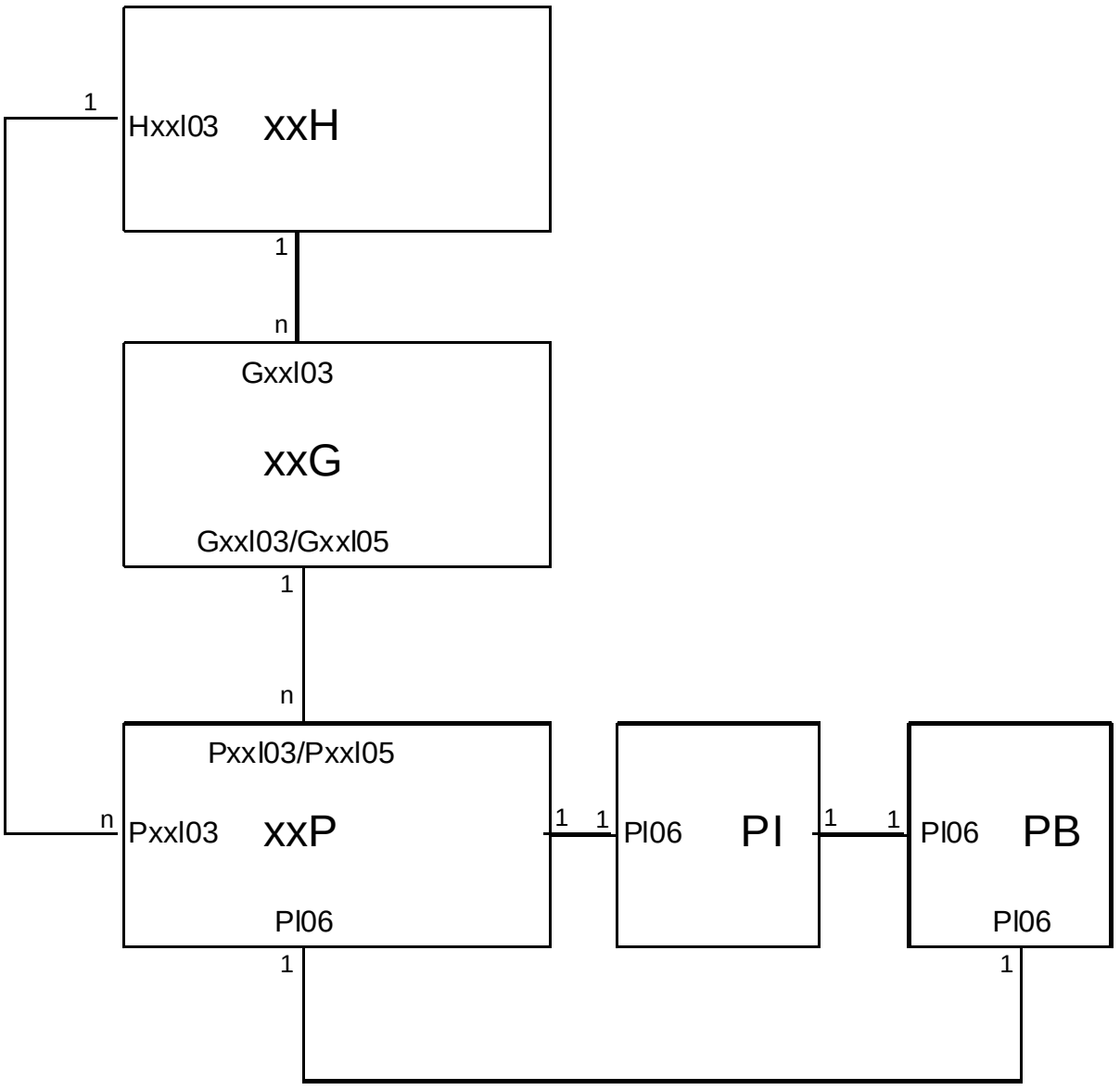


Figure 3: Predined Links between Household files to Individual data of Reference person and Spouse of Reference person

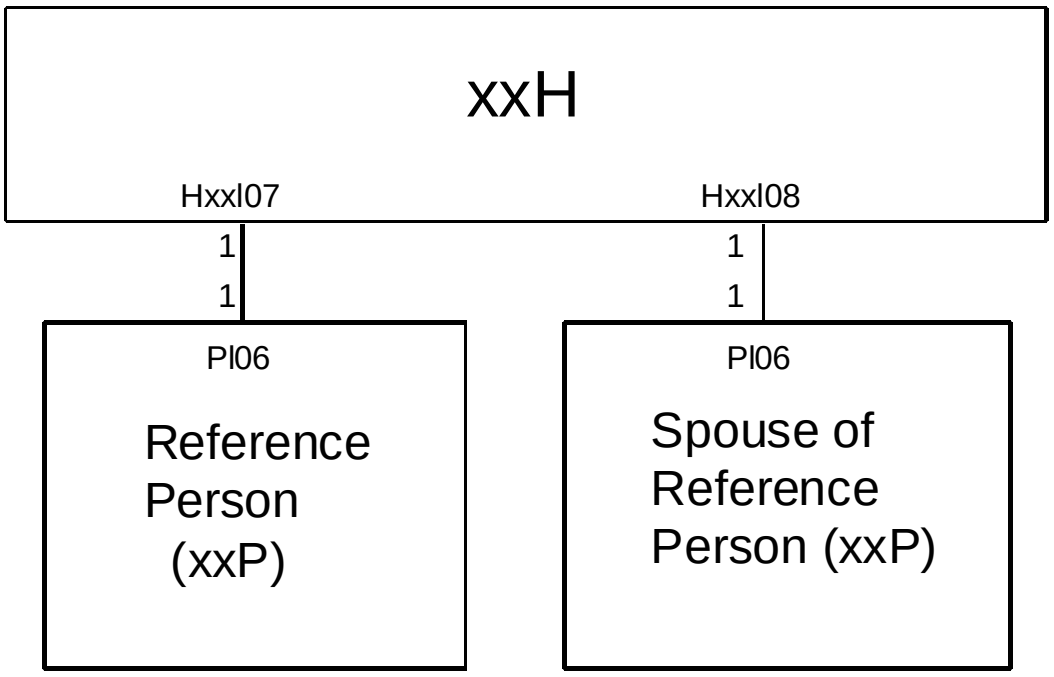


Figure 4: Predined links within the Person files to other family members

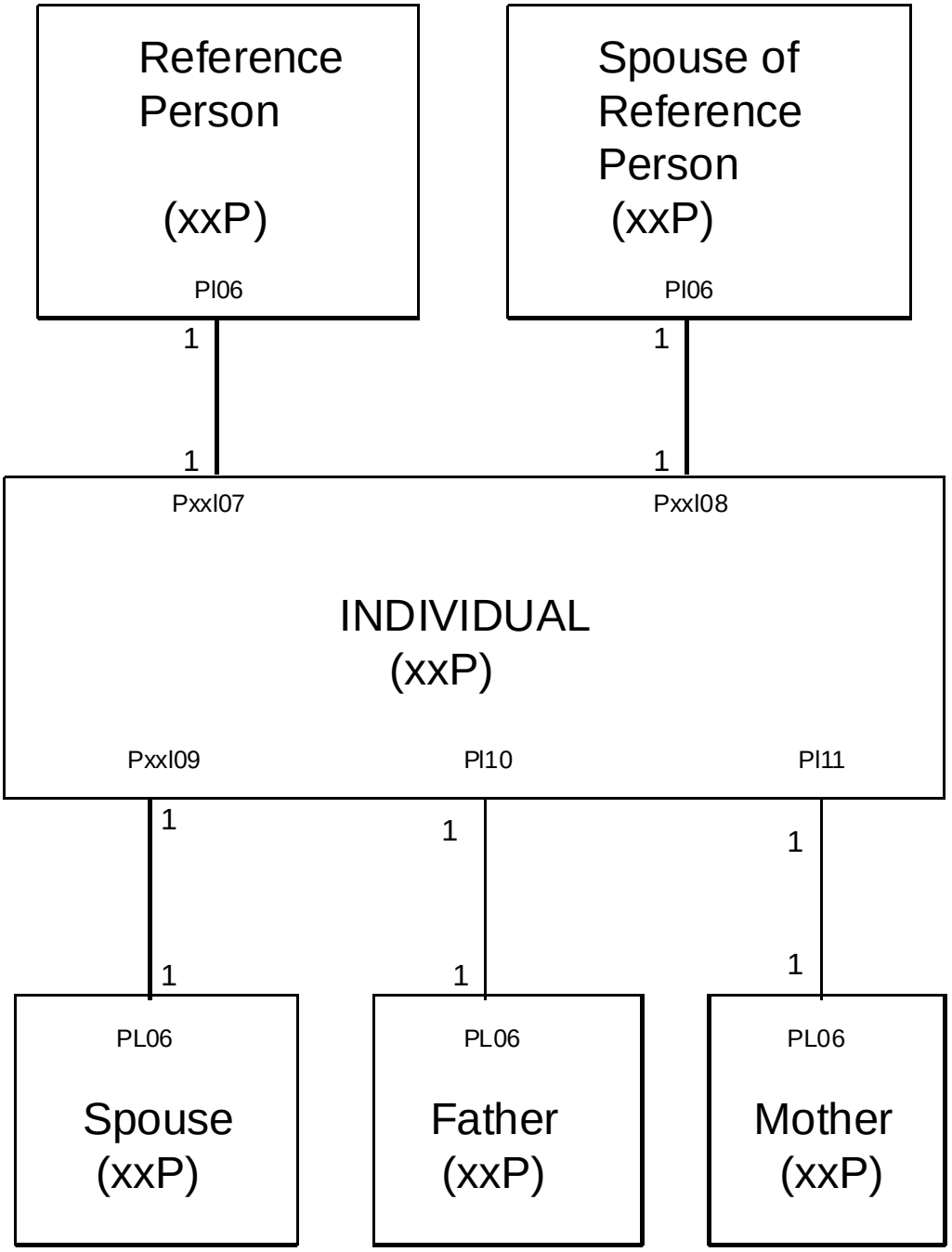


Figure 5: Longitudinal Match of Person files

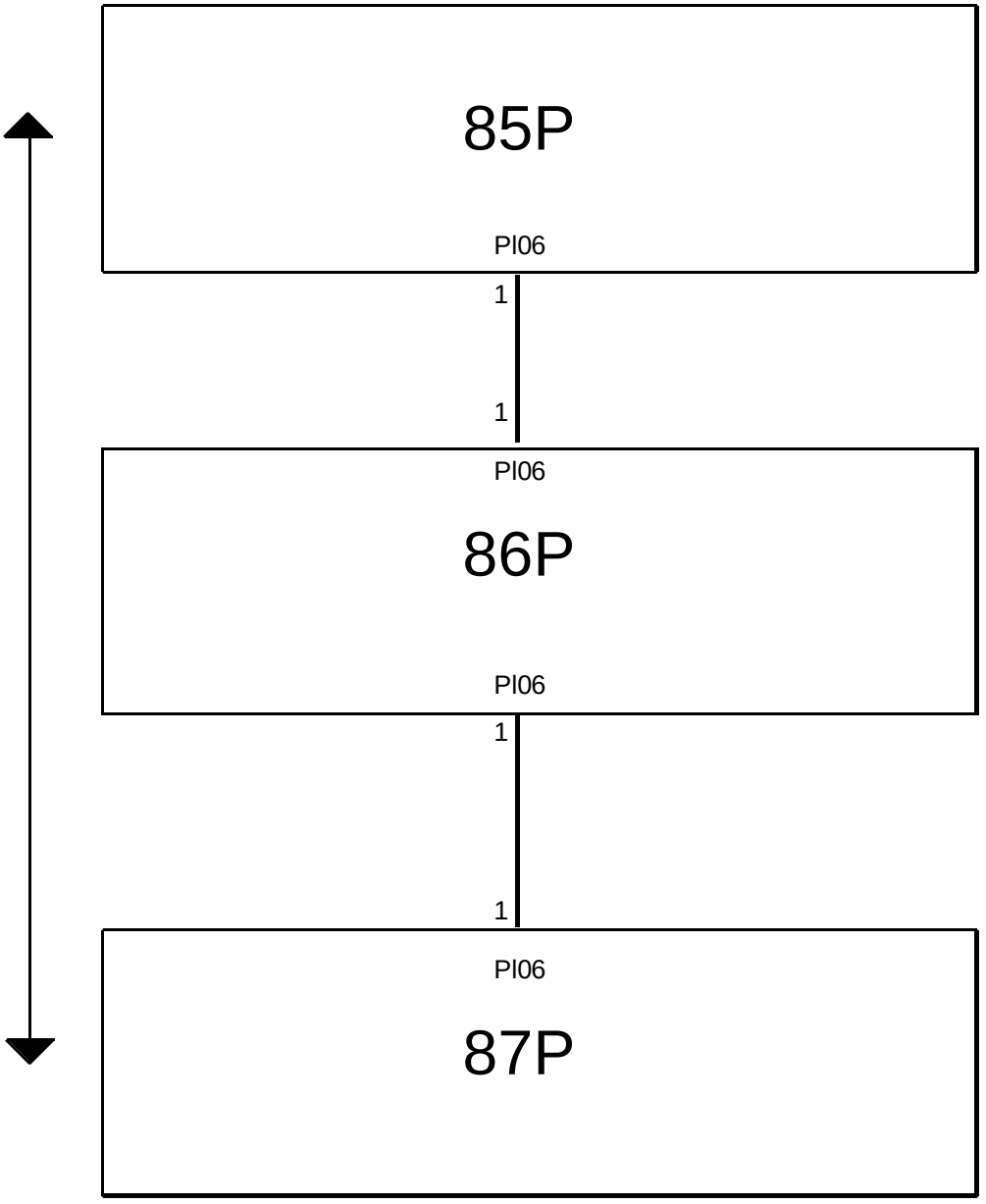


Figure 6: Longitudinal Match of Household Files from year t to year t+n (Forward Match)

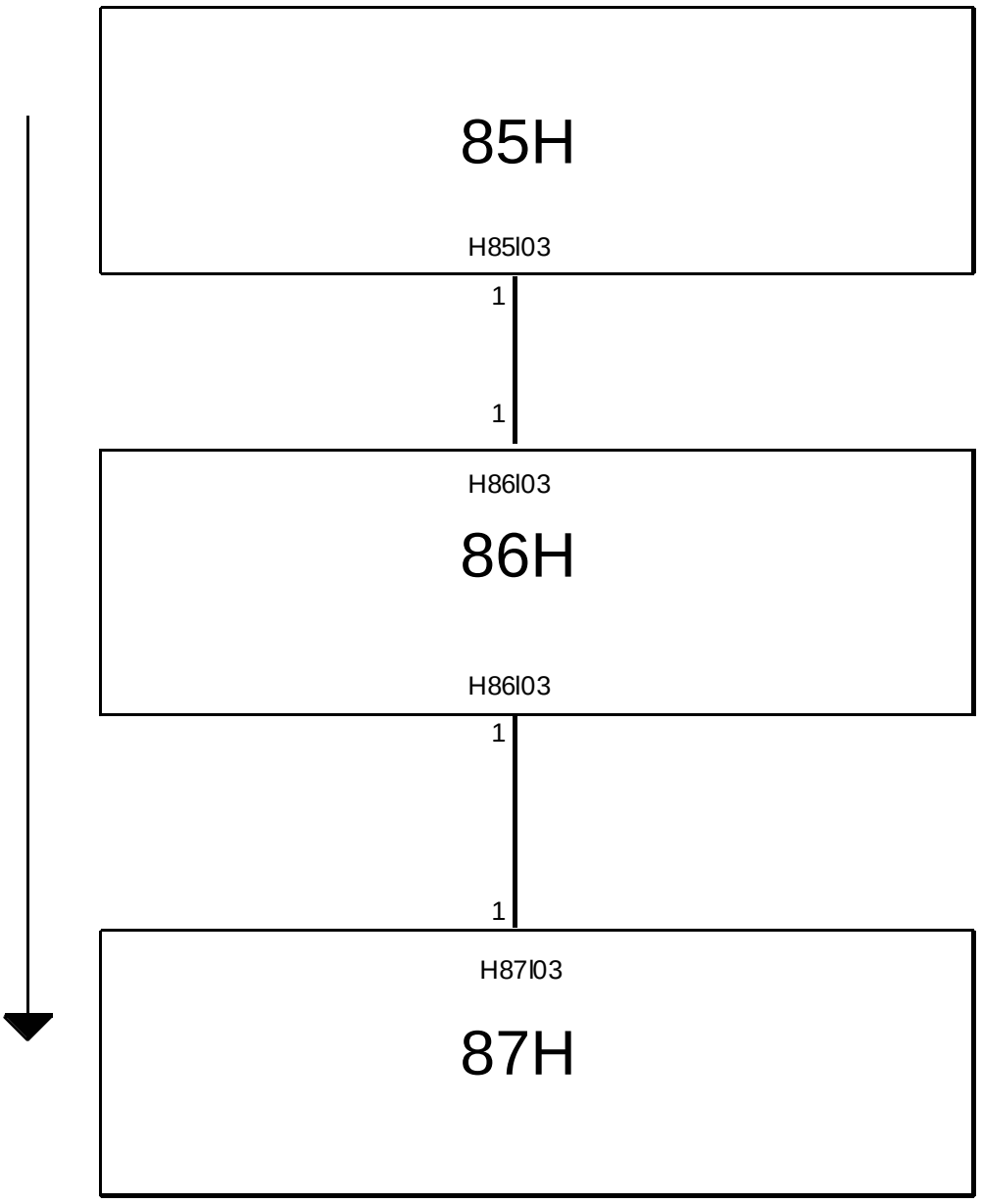


Figure 7: Longitudinal Match of Household Files from year t+n to year t (Backward Match)

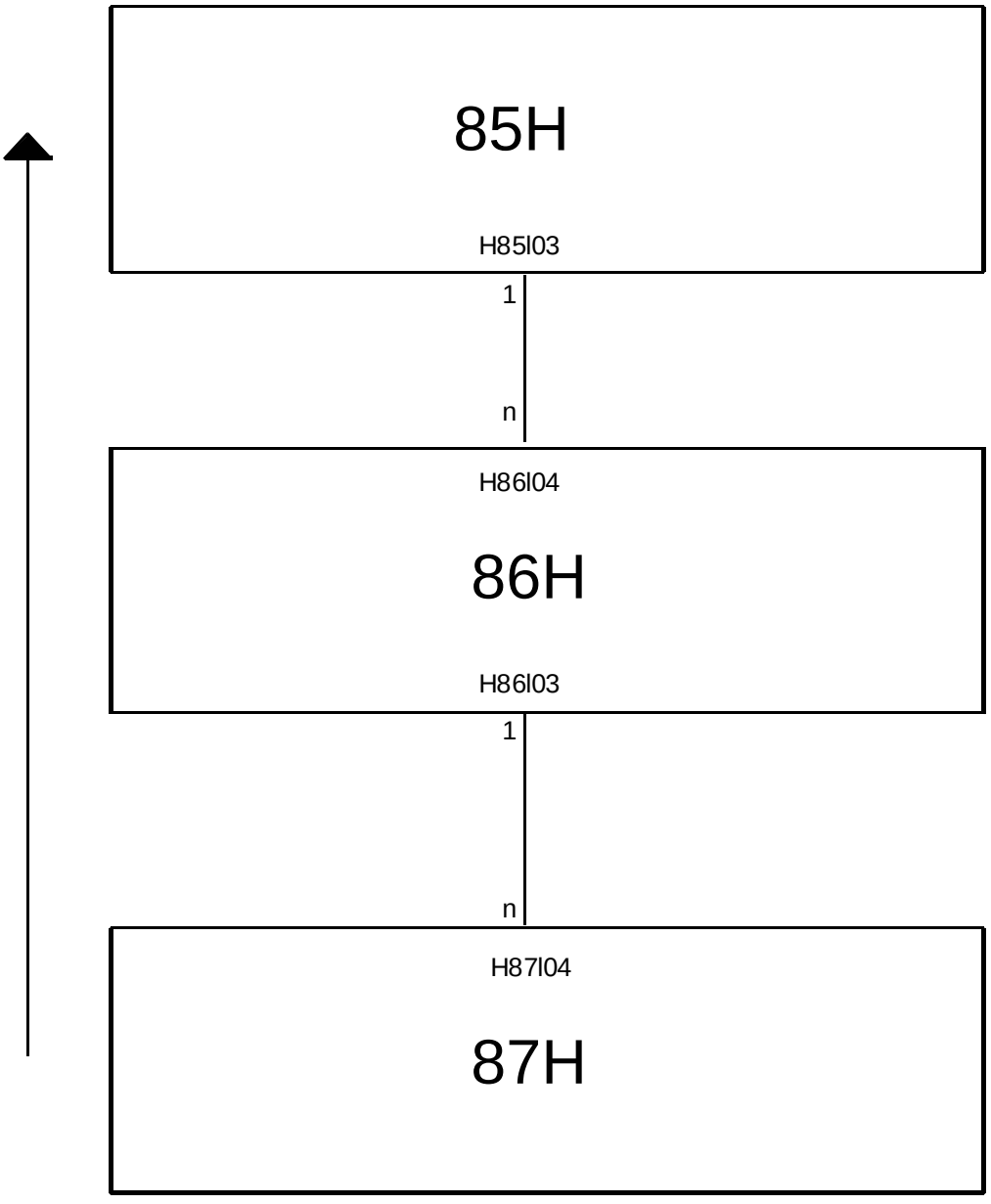


Figure 8: Longitudinal Match of Household files from years t+n to household of origin in year t

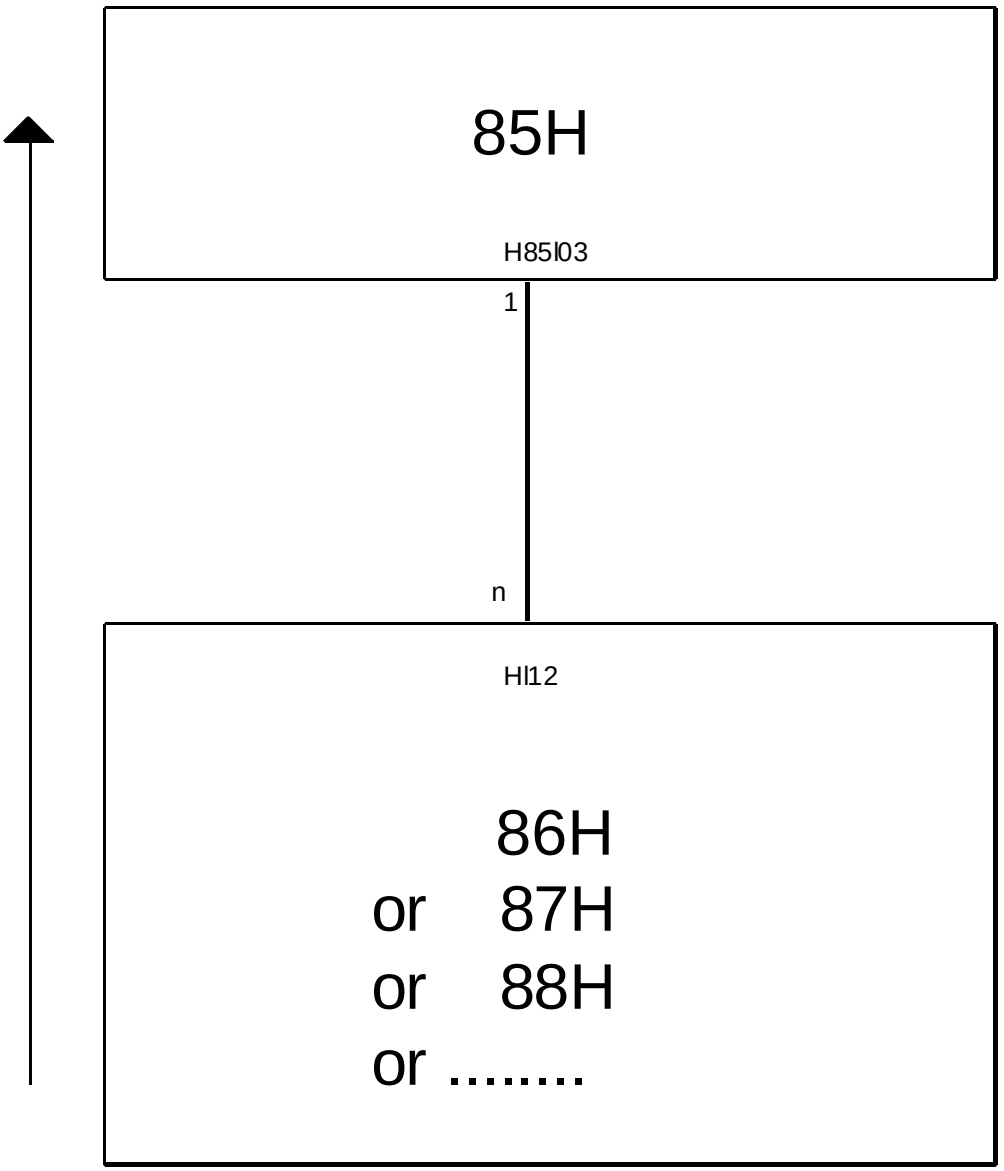
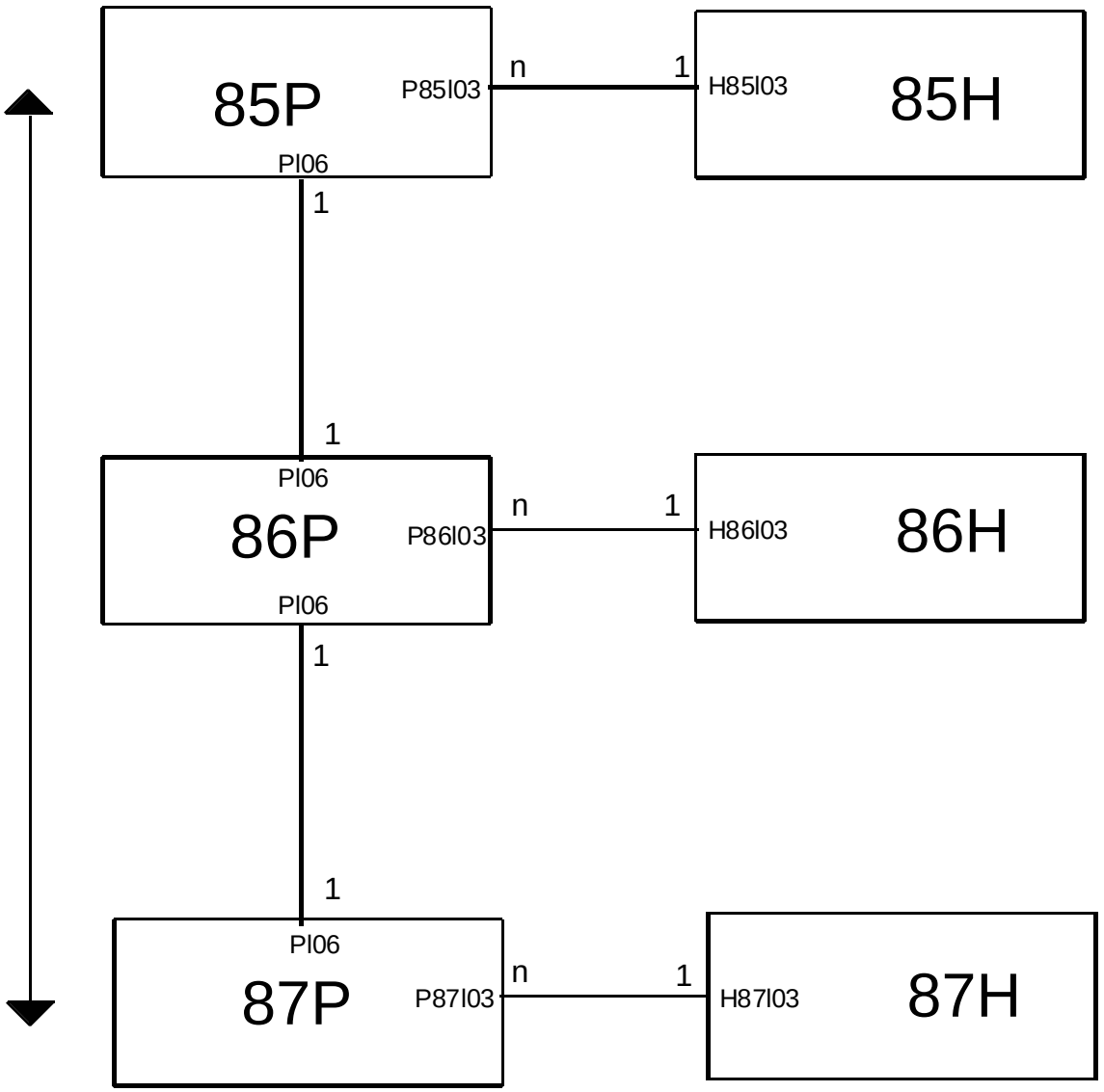


Figure 9: Longitudinal Match of Person files combined with a cross-sectional match between household and persons



B.6.4. Implementation of PACO Data Base

The relational data structure of the PACO Data Base is implemented as a set of system files for the statistical package SPSS for Windows. The (PACO) SPSS system files have a rectangular file structure, using 'columns' as variables and 'rows' as cases (observations). Therefore a SPSS system file is very similiar to the term 'Table" used in Data Base Management Systems. Each PACO system file contains data together with a dictionary. **All PACO Data Base files are supplied with a dictionary holding variable labels, value labels and missing value definitions.** Using the file management facilities of SPSS and the created key variables it is possible to combine multiple files or aggregate files into a single analysis file, not only for cross-sections but also for longitudinal populations and to combine data from different countries (see examples in section C).

B.7 Available PACO Files

a) Country files

Currently PACO data from six countries is available (see Table 4). We can learn from this figure that four of the PACO country data sets (France, Germany, Poland, Luxembourg and USA) have data from the 'eighties' and three countries (Hungary, Luxembourg, Hungary) have data from the 'nienties'.

Table 4: List of available countries and years in the PACO Data Base

	Reference year											
Country	83	84	85	86	87	88	89	90	91	92	93	94
France (Lorraine)			■	■	■	■	■	■				
Germany		■	■	■	■	■	■	■	■	■	■	■
Hungary										■	■	■
Luxembourg			■	■	■	■	■	■	■	■		
Poland					■	■	■	■				
UK									■	■	■	
USA	■	■	■	■	■							

The next Table 5 shows **that the main PACO files: (a) person inventory file, (b) household files and (c) person files for all countries have been created.** For three countries (France, Luxembourg and USA) files with information about income groups are available. **For two countries (Germany and UK) files holding the longitudinal Biographical vectors were created.**

The complete PACO Data Base (excluding Shadow files) currently contains 91 different system files and has a size of 210 MB (including Shadow files).

Table 5: List of available countries and files in the PACO DATA BASE					
	Person inventory file	Cross-sectional Files			Longitudinal Biographical Person file
		Household files	Group files	Person files	
Filename Country	PccPI	PccxxH	Pccxx G	Pccxx P	PccPB
France (Lorraine)	x	85-90	85-90	85-90	-
Germany	x	84-94	-	84-94	x
Hungary	x	92-94	-	92-94	-
Luxembourg	x	85-92	85-92	85-92	-
UK	x	91-93	-	91-93	x
USA	x	83-87	83-87	83-87	-

b) Variables

Due to non existence of some variables in some panels or years it was not possible to create for all countries and for all years the same number of PACO variables with non missing information. The section B.5.1 of the PACO User Gide contains a very detailed list of available PACO variables by country, year and unit.

C. The Matching of PACO Files

This section of the User Guide explains how the SPSS sytem files of the PACO database files can be combined and a selection of typical data structures can be created. This is done by giving example programs with SPSS syntax for following data structures:

C.1: Concatenation (combining) of three yearly cross-sectional files for one country

C.2: Concatenation (combining) of cross-sectional files from three countries for one identical year

C.3: Longitudinal match of household files from three contiguous years for one country

C.4: Longitudinal match of individual files from three contiguous years for one country

C.5: Longitudinal match of household files combining two countries and three years

C.6: Aggregation (transformation) of individual variables to household level for one country

C.7: Assigning (untransformed) variables of the reference person to a household file for one country

C.8: Spreading of untransformed variables from household file to individual file for one country

C.9: Matching individuals with information from partners by executing "Auto-join" for one country file

C.10: Longitudinal match of individual files combined with a cross-sectional match between household and individuals for one country

C.1 Concatenation (combining) of three yearly cross-sectional files for one country (Example 1)

Cross-sectional Household file concatenating (combining) cases from three years for one country			
File Structure: Household case file	Years hl02	Household key Hxxl03	Household Variables Hxx...
	85	1 "	hxx060 ...
	86	1 "	hxx060 ...
	87	1 "	hxx060 ...

SPSS Program: (Example 1)

```
add files
  /file=plu85h/rename= (h85l03=hxxl03) (h85l14=hxxl14)
            (h85060=hxx060) (h85i060=hxxi060)/in=h85
  /file=plu86h/rename= (h86l03=hxxl03) (h86l14=hxxl14)
            (h86060=hxx060) (h86i060=hxxi060)/in=h86
  /file=plu87h/rename= (h87l03=hxxl03) (h87l14=hxxl14)
            (h87060=hxx060) (h87i060=hxxi060)/in=h87
           /by hxxl03
           /keep=hxxl03 hl01 hl02 hxxl14 hxx060 hxxi060/map.
```

Comments:

procedure:	add files
input files:	cross-sectional household files
result file:	cross-sectional household case file containing different years
result variables:	household
renaming of variables:	necessary

C.2 Concatenation (combining) of cross-sectional files from three countries for one identical year (Example 2)

Cross-sectional Household file concatenating (combining) cases from three countries for one yea r			
File Structure: Household case file	Country hl01	Household key h85l03	Household Variables H85...
	France "	1 "	h85060 ...
	Germany "	1 "	h85060 ...
	Luxembourg "	1 "	h85060 ...

SPSS Program: (Example 2)

```
add files
  /file=pfr85h/in=hfr85
  /file=pge85h/in=hge85
  /file=plu85h/in=hlu85
          /keep = h85l03 hl01 hl02 h85l14 h85060 h85i060/map.
```

Comments:

procedure:	add files
input files:	cross-sectional household files
result file:	cross-sectional household case file containing different countries
result variables:	household
renaming of variables:	not necessary

C.3 Longitudinal match of household files from three contiguous years for one country (example 3)

Longitudinal Household file containing Household variables from one country				
File Structure: Household case file	Household Key H..l03	h85...	h86...	h87...
	1	h85060 ...	h86060 ...	h87060 ...
	2	h85060 ...	h86060 ...	h87060 ...
	n	h85060 ...	h86060 ...	h87060 ...

SPSS Program: (Example 3)

```
match files
    /file=plu85h/rename= (h85l03=hxxl03)/in=h85
    /file=plu86h/rename= (h86l03=hxxl03)/in=h86
    /file=plu87h/rename= (h87l03=hxxl03)/in=h87
            /by hxxl03
            /keep=hxxl03 hl01
             h85l14 h86l14 h87l14
             h85060 h86060 h87060
             h85i060 h86i060 h87i060/map.
```

Comments:

procedure:	match files
input files:	cross-sectional household files used as case files
result file:	longitudinal household case file
result variables:	household
keys:	primary keys: h85l03,h86l03,h87l03 for case files
reordering of variables:	recommended in order to be able to use the 'to' facility of SPSS
in subcommand:	recommended; creates variables which indicates separately for each year whether a case is available or not

The(household) case files are matched by using the primary keys 'h85l03','h86l03','h87l03' after having renamed them to hxxl03

C.4 Longitudinal match of individual files from three contiguous years for one country (Example 4)

Longitudinal Person file containing Person variables from one country				
File Structure: Person case file	Person Key PL06	p85...	p86...	p87...
	1	p85060 ...	p86060 ...	p87060 ...
	2	p85060 ...	p86060 ...	p87060 ...
	n	p85060 ...	p86060 ...	p87060 ...

SPSS Program: (Example 4)

```
match files
   /file=plu85p/in=p85
   /file=plu86p/in=p86
   /file=plu87p/in=p87
            /by pl06
            /keep=pl06 p85l03 p86l03 p87l03
                 pl01 p85l14 p86l14 p87l14
                 p85060 p86060 p87060
                 p85i060 p86i060 p87i060/map.
```

Comments:

procedure: match files
input files: cross-sectional individual files used as case files
result file: longitudinal individual case file
result variables: individual
keys: primary key: pl06 for case files
reordering of variables: recommended in order to be able to use the 'to' facility of SPSS
in subcommand: recommended; creates variables which indicates separately for each year whether a case is available or not

The (individual) case files are matched by using the primary key 'pl06'.

C.5 Longitudinal match of household files combining two countries and three years (example 5)

Longitudinal Household file containing Household variables from three countries					
File Structure: Household case file	Country hl01	Household Key Hxxl03	h85...	h86...	h87...
	France "	1 ...	h85060 ...	h86060 ...	h87060 ...
	Luxembourg "	1 ...	h85060 ...	h86060 ...	h87060 ...

SPSS Program: (Example 5)

```
match files
    /file=pfr85h/rename= (h85l03=hxxl03)/in=hfr85
    /file=pfr86h/rename= (h86l03=hxxl03)/in=hfr86
    /file=pfr87h/rename= (h87l03=hxxl03)/in=hfr87
    /file=plu85h/rename= (h85l03=hxxl03)/in=hlu85
    /file=plu86h/rename= (h86l03=hxxl03)/in=hlu86
    /file=plu87h/rename= (h87l03=hxxl03)/in=hlu87
            /by hl01 hxxl03
            /keep= hl01 hxxl03
                h85l14 h86l14 h87l14
                h85060 h86060 h87060
                h85i060 h86i060 h87i060/map.
```

Comments:

procedure:	match files
input files:	cross-sectional household files used as case files
result file:	longitudinal household case file containing two countries
result variables:	household
keys	primary keys: hl01/h85l03,hl01/h86l03 hl01/h87l03 for case files
reordering of variables:	recommended in order to be able to use the 'to' facility of SPSS
in subcommand:	recommended; creates variables which indicates separately for each year whether a case is available or not

The (household) case files are matched by using the primary keys
'hl01/h85l03','hl01/h86l03','hl01/h87l03'
after having renamed them to hxxl03

C.6 Aggregation (transformation) of individual variables to household level for one country (example 6)

Complex Cross-sectional file combining Household and aggregated Person variables to household level for one country			
File Structure: Household case file	Household Key Hxxl03	Household variables	Aggregated variables from Person File
	1	h85060	h85swork
	2	h85060	h85swork
	n	h85060	h85swork
Operation using	(Get)		Aggregation/ Table Match
Key in Household file	Hxxl03		Hxxl03
Key in Person file			p85l03

SPSS Program: (Example 6)

```
get file=plu85p/keep= p85l03 p85317/map.        /* Step 1 */

compute p85swork = 0.                          /* Step 2 */
if (p85317 eq 1) p85swork = 1.

SORT CASES BY p85l03.                           /* Step 3 */

aggregate outfile=*/break=p85l03                /* Step 4 */
 /h85swork = sum (p85swork).

match files  table=*                           /* Step 5 */
        /rename=(p85l03=h85l03)/in=p85
        /file=plu85h/in=h85
        /by h85l03
        /keep= h85l03 h85060 h85swork/map.
```

Comments:

procedure:	multi-step: get/sort/aggregate/match
input files:	cross-sectional individual file used as start/aggregation/table file
	cross-sectional household file used as case file
result file:	cross-sectional household case file
result variables:	household/individual aggregation variable
keys:	secondary key 'p85l03' from start file
	primary key 'h85l03' from case file
Step 1	get variables from the individual start file 'plu85p' and
	create an active file which will be aggregated
Step 2:	prepare the aggregation variable p85swork
Step 3:	sort the (individual)active file by secondary key 'p85l03'
Step 4:	aggregate the active file with the break variable 'p85l03'
	using the summary function sum and create a new active file (*)
Step 5:	match the (household) case file 'plu85h' with the aggregation
	file used as table file using the primary key 'h85l03' for
	the case file and the secondary key 'p85l03' for the table file
	after having renamed 'p85l03' to 'h85l03'

C.7 Assigning (untransformed) variables of the reference person to a household file for one country (example 7)

Complex Cross-sectional file combining Household and variables from reference person on household level for one country			
File Structure: Household case file	Household Key Hxxl03	Household variables	Reference Person variables
	1	h85060	r85204
	2	h85060	r85204
	n	h85060	r85204
Operation using	(Get)		Table Match
Key in Household file	Hxxl03		Hxxl07
Key in Person file			pl06

SPSS Program: (Example 7)

```
get file = plu85h/rename=(h85l07=merge)                    /* Step 1 */
                 /keep=h85l03 merge h85060.
sort cases by merge.                                       /* Step 2 */

match files table=plu85p                                   /* Step 3 */
           /rename=(pl06=merge)(p85204=r85204)/in=p85
           /file=*/in=h85
           /by merge
           /keep= h85l03 merge h85060 r85204.
rename variables (merge=h85l07).                           /* Step 4 */
```

Comments:

procedure:	multi step: get/sort/table match
input file:	cross-sectional individual file used as table file
	cross-sectional houshold file used as start/case file
result file:	cross-sectional household case file
result variables:	household/reference person
keys:	primary key 'pl06' for table file
	secondary key 'h85l07' for case file
renaming of variables:	not necessary for variables from table file, but recommended
Step 1 :	read variables from the (household) start file 'plu85h' and
	rename the secondary key 'h85l07' (ID-reference person)
	to 'merge' and create the active file which acts
	as case file for the match
Step 2:	sort the (household) case file by the renamed secondary key 'merge'
Step 3:	match the household case file (*) with the individual table
	file 'plu85p' using the the secondary key 'merge' for the
	case file and the primary key 'pl06' for the table file
	after having renamed
	a) the id variable 'pl06' to 'merge'
	b) the reference person variable 'p85204' into 'r85204'.
Step 4:	rename the key variable 'merge' into the old variable name 'h85l07'

C.8 Spreading of untransformed variables from household file to individual file for one country (example 8)

Complex Cross-sectional Person File combining variables from the Household and the Person file for one year			
File Structure: Person case file	Person key pl06	Person variables Pxx...	Household variables Hxx...
	1	p85060	h85060
	2	p85060	h85060
	n	p85060	h85060
Operation using	(get)		Table Match
Key in Person file		pl06	pxxl03
Key in Household file			h85l03

SPSS Program: (Example 8)

```
get file=plu85p/keep=pl06 p85l03 pl01 p85060/map.          /* Step 1 */

SORT CASES BY p85l03.                                     /* Step 2 */

MATCH FILES                                               /* Step 3 */
    /TABLE=plu85h/RENAME=( h85l03 = p85l03)/IN= hh85
    /FILE=*/IN= pp85
    /BY p85l03
    /keep=pl06 p85l03 hl01 h85060 p85060 /map.
```

Comments:

procedure:	get/sort/table match
input files:	cross-sectional household file used as table file
	cross-sectional individual file used as start/case file
result file:	cross-sectional individual case file
keys:	primary key 'h85l03' for table file
	secondary key 'p85l03' for case file
variables:	household/individual
Step 1:	get variables from the (individual) start file 'plu85p' and produce an active file which is used as case file
Step 2:	sort the (individual) case file by the secondary key 'p85l03'
Step 3:	match the (individual) case file (*) with the (household) table file 'plu85h' using the secondary key 'p85l03' for the case file and the primary key 'h85l03' for the table file after having renamed the key variable 'h85l03' into 'p85l03'

C.9 Matching individuals with information from partners by executing "Auto-join" for one country file (example 9)

Complex Cross-sectional Person File combining variables from the Person and his/her Spouse for one year			
File Structure: Person case file	Person key pl06	Person variables Pxx...	Spouse Variables
	1	p85060	c85060
	2	p85060	c85060
	n	p85060'	c85060
Operation using	(get)		Auto-Join
Key in Person file		pl06	pxxl08

SPSS Program: (Example 9)

```
get file = plu85p/rename=(p85l09=merge)                    /* Step 1 */
     /keep=pl06 merge p85060.
sort cases by merge.                                      /* Step 2 */

match files table=plu85p                                  /* Step 3 */
           /rename=(pl06=merge)(p85060=c85060)/in=i85
           /file=*/in=p85
           /by merge
           /keep= pl06 merge p85060 c85060.
rename variables (merge=p85l09).                          /* Step 4 */
```

Comments:

procedure:	multi step: get/sort/table match (auto-join)
input file:	cross-sectional individual file used as start/case/table file
result file:	cross-sectional individual case file
result variables:	individual/partner
keys:	primary key 'pl06' for table file secondary key 'p85l09' for case file
renaming of variables:	necessary for variables from table file
Step 1 :	read variables from the (individual) start file 'plu85p' and rename the secondary key 'p85l09'(partner-id) to 'merge' and create the active file which acts as case file for the match
Step 2	sort the (individual) case file by the renamed secondary key 'merge'
Step 3:	match the (individual) case file (*) with the (individual) table file 'plu85p' by using the secondary key 'merge' for the case file and the primary key 'pl06' for the table file after having renamed: a) the key variable 'pl06' to 'merge' b) the partner variable 'p85060' into 'c85060'.
Step 4:	rename the variable 'merge' into the old variable name 'p85l09'

C.10 Longitudinal match of individual files combined with a cross-sectional match between household and individuals for one country (example 10)

Complex Longitudinal Person file combining household and person variables for one country								
File Structure: Person case file	Person Key Pl06	Household Key Hxxl03	p85...	h85...	p86...	h86...	p87...	h87...
	1	1	p85317	h85060	p86317	h86060	p87517	h87060
	2	3	p85317	h85060	p86317	h86060	p87517	h87060
	n	n	p85317	h85060	p86317	h86060	p87517	h87060

SPSS Program: (Example 10)

```
match files                                                    /* Step 1 */
     /file=plu85p/in=p85
     /file=plu86p/in=p86
     /file=plu87p/in=p87
                /by pl06
                /keep=pl06 p85l03 p86l03 p87l03 pl01
                      p85317 p86317 p87317/map.

SORT CASES BY p85l03.                                          /* Step 2 */

MATCH FILES                                                    /* Step 3 */
      /TABLE= plu85h/RENAME=( h85l03 = p85l03)/IN= hh85
      /FILE=*/IN= pp85
      /BY p85l03
      /keep= hl01
             h85060 h85i060
             pl06 p85l03 p86l03 p87l03 pl01
             p85317 p86317 p87317
      /map.

SORT CASES BY p86l03.                                          /* Step 4 */

MATCH FILES                                                    /* Step 5 */
      /TABLE= plu86h/RENAME=( h86l03 = p86l03)/IN= hh86
      /FILE=*/IN= pp86
      /BY p86l03
      /keep= hl01
             h85060 h85i060
             h86060 h86i060
             pl06 p85l03 p86l03 p87l03 pl01
             p85317 p86317 p87317
            /map.

SORT CASES BY p87l03.                                          /* Step 6 */

MATCH FILES                                                    /* Step 7 */
     /TABLE= plu87h/RENAME=( h87l03 = p87l03)/IN= hh87
```

```
/FILE=*/IN= pp87
/BY p87l03
/keep= hl01
       h85060 h85i060
       h86060 h86i060
       h87060 h87i060
       pl06 p85l03 p86l03 p87l03 pl01
       p85317 p86317 p87317
      /map.
```

Comments:

procedure:	mukti step: match files/sort/table match
input files:	cross-sectional household files used as table files
	cross-sectional individual files used as case file
result file:	logitudinal individual case file
result variables:	household/individual
keys:	primary key: pl06 for step 1
	primary key: h85l03,h86l03,h87l03 for step 2-7
	secondary key: p85l03,p86l03,p87l03 for step 3,5,7
reordering of variables:	recommended in order to be able to use the 'to' facility of SPSS
in subcommand:	recommended; creates variables which flags whether a case is available in all years
Step 1:	longitudinal file match of person files 'plu85p','plu86p','plu87p'
	using primary key 'pl06'
Step 2:	sort the (individual)case file by secondary key 'p85l03'
Step 3:	table match between first household file 'plu85h' and case file
Step 4:	sort the(individual) case file by secondary key 'p86l03'
Step 5:	table match between second household file 'plu86h' and case file
Step 6:	sort the (individual)case file by secondary key 'p87l03'
Step 7:	table match between third household file 'plu87h' and case file

D PACO Publication List
(November. 1995)

Joachim Frick, Irena Topinska ,Gert G. Wagner, and Klaus Mueller (1993): Income Inequality and Poverty Dynamics in Poland and East-Germany before and under Transition
PACO Research Paper, Document no 6

G. Ghellini, N. Pannuzi, L. Stanghellini (1995):
Deprivation Pattern in the USA
PACO Research Paper, Document no 10

Marlis Riebschläger (1994):
La variance des estimateurs d'un panel ménage - la méthode des groupes aléatoires appliquée au panel luxembourgeois. Document PSELL no 65.

Marlis Riebschläger (1995):
A review of weighting methods employed by panel studies included in the PACO project.
Proceedings : XV^{e} Jounées de l'Association d'Economie Sociale. L'Analyse Longitudinale en Economie Sociale
and
PACO Research Paper, Document no 7

Gaston Schaber (1993):
Developing Comparative Databases
Research Paper, CEPS/INSTEAD Luxembourg 1993

Gaston Schaber and **Günther Schmaus** (1993):
Problems of Aging
Working Papers of the ESF Network on Household Panel Studies: Paper Number 86

Gaston Schaber, Günther Schmaus and **Marlis Riebschläger**(1994):
Looking at Intergenerational Relations in Longitudinal Panel Studies on Individual and Households.
Research Paper, CEPS/INSTEAD Luxembourg 1994

Gaston Schaber Günther Schmaus and **Gert G. Wagner** (1992):
Building up an international comparative panel database: The PACO-Project
in: Proceedings of Statistics Canada Symposium 92, Design and Analysis of Longitudinal Surveys (August 1993 Statistics Canada)
and ESF Working Paper Number 29, University of Essex 1992

Gaston Schaber, Günther Schmaus and **Gert G. Wagner** (1993):
The PACO Project
PACO Research Paper, Document no 1

Gaston Schaber, Günther Schmaus and **Gert G. Wagner**(1993):
Building up a Cross-National Comparative Household Panel Database
Research Paper CEPS/INSTEAD 1991(revised 1993)

Gaston Schaber, Günther Schmaus, Gert G. Wagner and (1993):
Making Data European - The PACO Approach for Household Panel Studies
-for Making Data European Conference in Luxembourg: 15/16/17 April 1993
Research Paper CEPS/INSTEAD Luxembourg 1993

Günther Schmaus (1994):
Technical Specifications of the PACO Database
Research Paper CEPS/INSTEAD Luxembourg 1994

Günther Schmaus and **Marlis Riebschläger**(1994):
Variable Specifications for the PACO Database
Research Paper CEPS/INSTEAD Luxembourg 1994

Günther Schmaus and **Marlis Riebschläger**(1995):
PACO User Guide
Paco Research Paper, Document no 13 CEPS/INSTEAD Luxembourg 1995

Günther Schmaus and **Gaston Schaber** (1994):
Pattern of Retirement and Exiting Out of Work
Research Paper CEPS/INSTEAD Luxembourg 1994

Charanjit Singh (1995):
A comparative Analsis of Attrition in Household Panel Studies
Paco Research Paper, Document no 14

Marcia Taylor (1995):
New Possibilities for Comparative Research
PACO Research Paper, Document no 8

Marcia Taylor and **Gaston Schaber** (1995):
Data Confidentiality
PACO Research Paper, Document no 9

MarciaTaylor, Günther Schmaus and **Gert G. Wagner** (1993):
How to build a user friendly Household Panel Data Base
Working Papers of the ESF Network on Household Panel Studies: Paper Number 62

E.2 Macro Economic Variables

In addition to the micro-data a set of macro-economic indicators is provided together with variables that indicate the country and the year, so that these variables can be linked easily to the micro-data.
The most recent information has been taken from the following sources:

- EUROSTAT CD : Electronic statistical yearbook of the European Community.Edition 2 / 1993
- EUROSTAT : Basic Statistics of the Community - Comparison with the principle partners of the Community
- Main Economic Indicators - Historical Statistics: Prices, Labour and Wages 1962 - 1991.
  OECD, Paris 1993.
- National Accounts - Main Aggregates 1960 - 1991.
  OECD, Paris 1993.

Variable-name	Label	Description	Source
m1	population total (*1000)		Eurostat CD 2.2.1.2
m2	population : men (*1000)		Eurostat CD 2.2.1.2
m3	population : women (*1000)		Eurostat CD 2.2.1.2
m4	% older 64, men	percentage of 65 or older of total population, yearly average	Eurostat Basic Statistics 3.11
m5	%older64,women	percentage of 65 or older of total population, yearly average	Eurostat Basic Statistics 3.11
m6	foreigners men and women (*1000)		Eurostat CD 2.2.1.5.1
m7	foreigners men (*1000)		Eurostat CD 2.2.1.5.1

Variable-name	Label	Description	Source
m8	foreigners women (*1000)		Eurostat CD 2.2.1.5.1
m9	working pop tot. men and women (*1000)	working population total , men and women	Eurostat CD 2.3.1.1 (see expl. below)
m10	working pop tot. men (*1000)	working population total , men	Eurostat CD 2.3.1.1 (see expl. below)
m11	working pop tot. women (*1000)	working population total , women	Eurostat CD 2.3.1.1 (see expl. below)
m12	working pop civ. men and women (*1000)	working population civilian, men and women	Eurostat CD 2.3.1.2 (see expl. below)
m13	working pop civ. men (*1000)	working population total , men	Eurostat CD 2.3.1.2 (see expl. below)
m14	working pop civ. women (*1000)	working population total , women	Eurostat CD 2.3.1.2 (see expl. below)
m15	stand. unempl. rate %	standardized unemployment rate percentage of total labour force seasonnally adjusted annual averages	OECD : Main Economic Indicators
m16	activity rate total men / women		Eurostat CD 2.3.1.6.1 (see expl. below)
m17	activity rate total men		Eurostat CD 2.3.1.6.1 (see expl. below)
m18	activity rate total women		Eurostat CD 2.3.1.6.1 (see expl. below)
m19	activity rate civ. men / women	activity rate civilian , men and women	Eurostat CD 2.3.1.6.2 (see expl. below)

Variable-name	Label	Description	Source
m20	activity rate civ. men	activity rate civilian , men	Eurostat CD 2.3.1.6.2 (see expl. below)
m21	activity rate civ. women	activity rate civilian , women	Eurostat CD 2.3.1.6.2 (see expl. below)
m22	hourly earn. nace1-5 men / women	average hourly earnings men and women manual workers in industry (NACE 1-5) unit : ECU	Eurostat CD 2.6.1.1 (see expl. below)
m23	hourly earn. nace1-5 men	average hourly earnings men manual workers in industry (NACE 1-5) unit : ECU	Eurostat CD 2.6.1.1 (see expl. below)
m24	hourly earn. nace1-5 women	average hourly earnings women manual workers in industry (NACE 1-5) 2nd value of the year (out of 2) unit : ECU	Eurostat CD 2.6.1.1 (see expl. below)
m25	consumer price index	consumer price index 1985=100 FRANCE: all items GER : all items, excluding seasonal items LUX : all items UK : all items, excluding seasonal items USA : all items	OECD : Main Economic Indicators
m26	GDP at curr. prices/PPPs (bill.$)	Gross domestic product at current prices and current PPPs unit : billions US Dollars	OECD : National Accounts

Variable-name	Label	Description	Source
m27	GDP per head at curr. prices/PPPs ($)	Gross domestic product per head at current prices and current PPPs unit : US Dollars	OECD : National Accounts
m28	curr. expendit. soc.protect, % of GDP	Current expenditure on social protection as percentage of gross domestic product at market prices	Eurostat Basic Statistics 3.31
m29	soc.protect.benefits /head (ECU)	Social protection benefits per head Total population yearly averages unit : ECU	Eurostat Basic Statistics 3.32
m30	soc.protect.benefits/head (PPPs)	Social protection benefits per head Total population yearly averages unit : PPS	Eurostat Basic Statistics 3.33
m31	exch. rate average value of period (ECU)	exchange rate average value of period (ECU)	Eurostat CD 1.3.3.1 (see expl. below)
m32	exch. rate value at end of period (ECU)	exchange rate value at the end of the period (ECU)	Eurostat CD 1.3.3.1 (see expl. below)
m33	currency		Eurostat CD 1.3.3.1 (see expl. below)

The numbers added to the Eurostat Basic Statistics refer to the related numbering system.
Those added to the EUROSTAT CD information refer to the EUROSTAT information and indicator system.

Explanations from Eurostat CD
The following text has been extracted from the EUROSTAT CD.

2.3.1.1. WORKING POPULATION TOTAL
2.3.1.6.1. ACTIVITY RATE TOTAL
2.3.1.6.2. ACTIVITY RATE CIVILIAN

National Information
1 total working population
The total working population comprises all those persons who were employed during the reference period as well as those out of work, i.e persons who would have wished to work during this period if they had had the opportunity. The civilian working population comprises these same persons, excluding the armed forces.

716 Population - Inhabitants
Definition :
On a given date the total population of a country consists of all persons, national or foreign, who are permanentey settled in that country, even if they are temporarly absent from it (esa : resident population). This definition tallies with that used in the "1981 community census of population programme" (doc 3065/76).
Two types of population figures are to be distinguished: the total population (situation on january 1st) and the yearly mean of population. It's the yearly mean of total population that is considered as relational value for per in habitants (per head) calculations. In most countries, annual population estimates are based on the most recent census, with corrections for natural increase (births minus deaths) and net migration (immigrants minus emigrants). Although extremely accurate data are available on births and deaths, those on migrations are generally less reliable and thus estimates of total population may not be quite accurate. The member states produce their main estimate of population by sex and age as at 1 january (or 31 december) each year. Exceptions are the united kingdom, where the main estimate relates to 30 june and ireland, where it refers to mid-april. The statistical offices of these two countries have, however, provided eurostat with special estimates as at 1 january. The series cover employment and unemployment as well as strikes and are expressed in thousands

723 activity rate
The figures of working population are related to the total population figures.

Regional information

014/community labour force sample survey
The results of the survey refer exclusively to private households, since persons living in institutional households make up only a small fraction of the population (the community average is around 2%). The community survey is carried out in spring, but the precise period during which it takes place varies somewhat from one country to another. As the survey is conducted on a sample basis, results relating to small numbers of persons must be treated with caution. Great care must be taken when comparing the results with those of earlier surveys. This is mainly because the sample and the basis for grossing up the results may change from one survey to the next.

In addition, the community coding system has been slightly modified in order to increase the precision of the results and certain countries have modified their national questionnaires.
From 1983 onwards, the definitions are in conformity with the international labour office (ilo) recommendations, i.e.:
- unemployment includes people who have no job, are looking for a job, who have made serious efforts to find one, and who are immediately available for work.
- the labour force comprises persons in employment and the unemployed.

012/working population
The working population is calculated by adding the estimated number of unemployed to the number of persons in work at the time of the lfs.the working population is regionalized on the basis of the results of the lfs up to nuts level ii. For level iii the structure supplied by the member states is used. In france, working population for provence-alpes-côte d'azur and corse is corrected using the same structure as the employment figures supplied by insee.

1.3.3.1. Exchange rate:

51 ecu (european currency unit)
The ecu was introduced together with the european monetary system on 13 march 1979. It is composed of the same amounts of national currencies as the eua (european unit of account) which was formerly used for soec statistics.

These amounts are as follows :

1 ecu =

0,719	German Mark
1,31	French Francs
0,0878	Pound Sterling
140	Italian Lire
0,256	Dutch Florin
3,71	Belgian Francs
0,140	Luxembourg Franc
0,219	Danish Krone
0,00871	Irish Pound
1,15	Drachma

The equivalent of the ECU (and the EUA) in any currency is equal to the value of these amounts in that currency. The calculation of the ECU (and the EUA) equivalent is done via the us dollar, which has been chosen as giving the most representative currency all financial centres. The us-dollar exchange rates are communicated by the national bank of belgium to the commission, which uses them to calculate an ecu equivalent first in dollars and then in the currencies of the member states.
The daily equivalents in the different currencies calculated by the commission are published correct to six significant figures in the edition of the official journal of the european communities.

633 exchange rates
The rates of exchange stored in annual series are the annual average of the daily rates per country. These series are used to convert the values of all the aggregates provided by the invidual countries from national currency to ecu.

611 ecu exchange rate
For the calculation of the ecu exchange rate (see info 51). Monthly, quarterly and annual averages are weighted by the number of working days in the relevant period. Index figures are based on the average rate of one period.

2.6.1.1. Average gross hourly earnings of male and female manual workers in industry

National information

735 gross earnings
Gross earnings cover remuneration in cash paid directly and regularly by the employer ar the time of each wage payment, before tax deductions and social security contributions which are payable by the wage-earners and retained by the employer, and before fines. Payments for leave, public holidays, and other paid individual absences, are included in principe, in so far as the corresponding days or hours are also taken into account to calculate earnings per units of time.

734 non-manual workers
All salaried persons, who are not included in the definition of manual workers are considered to be non-manual. The group thus includes not only non-manual workers in the strict sense, but also foremen, overseers and administrative, technical and commercial staff (directors and mangers with similar broad reponsibilities in the running of the enterprises were excluded).

F National Documentation

F.1 Description of Panel Studies

F.1.1 France

Panelname : ESEML
Enquête Socio-Economique auprès des Ménages Lorrains
or
Panel des Ménages Lorrains

Host Institutes : ADEPS
Equipe de recherche en Analyse Dynamique des Effets des Politiques Sociales
URA CNRS n° 1167 Université Nancy 2
and
Direction Régionale en Lorraine de l'Institut National de la Statistique et des Etudes Economiques (INSEE)

Sponsors : Institut National de la Statistique et des Etudes Economiques (INSEE)
Ministère de l'Education Nationale (MEN)
Centre National de le Recherche Scientifique (CNRS)
Ministère de la Recherche et de la Technologie (MRT)
Commissariat Général du Plan (CGP)
Conseil Régional de Lorraine
Commission des Communautés Européennes (DG V et FSE)
Caisse Nationale des Allocations Familiales (CNAF)
Ministère de l'Equipement, du Logement, de l'Aménagement du Territoire et des Transports (MELATT)
Mission Interministérielle Recherche Expérimentation (MIRE)
des Ministères des Affaires Sociales et du Travail

User Support : ADEPS
4, rue de la Ravinelle
co 26
54035 NANCY cedex
FRANCE
Tel. (33) 83 30 58 41
Fax (33) 83 35 83 90
e-mail : jeandid@droit-eco.u-nancy.fr

Availability : The data are accessible in the labatory ADEPS (Nancy) on Unix station or on PC after transfer of the data. The accessibility is decided individually.

Anonymization : The data are completely anonymized (no name and no residence code).

No of Waves
currently available: 6

Years covered : 1985 to 1990 (the first wave 1985 is limited to a subsample of about 700 households).

Time Span : The field work campaign starts every year in november ending in december. The reference period for the interview is the twelve months subsequent to when the survey started.

Geographic Universe : Region of Lorraine (France).

Reference Population : Anyone living in Lorraine, except the persons living in a collective household (for example, in an old people's home).

Case units : Households
Income groups within a household
Individuals

Household : a household consists of all persons who live together in a dwelling unit (house, appartment, group of rooms or single room). Persons within a household can be related to each other or not.

Income group : in a household in which several persons have individual income, different economic arrangements are possible. An Income group is a group of persons within a household who constitute an economic unit because they share their incomes. The concept is put into effect in accordance with strict rules.

Questionnaires : Questionnaire A - Household (one per household)
Questionnaire B - Income group (one per Income group located within the household)
Questionnaire C and D : individual (one questionnaire per household member). The questionnaire C concerns the members below the age of 16, and/or those who are in full-time education and have never interrupted their school education for more than one year.
In 1986 and 1987, there is a thematic questionnaire F concerning the economic behaviours after a large decrease of income.
From the 1988 wave to 1990 wave, the part of the

questionnaire D concerning the individual biography is detached in a specific questionnaire G.
In 1989 and 1990 the questionnaire B is included in the A and the questionnaire C is included in the D.

Method of Data Collection: Face to face interview.

Sampling Procedure : The initial sample is a simple random sample of persons drawn from the Echantillon Démographique Permanent (EDP) de l'INSEE. Each person leads to one household. Every persons who live in this household are interview and constitute the intial sample (persons who are followed in the successive waves). In 1988 and 1990 extentions were added to the initial sample by drawing, in the EDP, persons born after 1985.
For further information, see reference 2.

Use of weights : It's possible to use a weight variable which ajustes the initial sampling procedure. This variable don't take account for attrition. A second weight variable including the correction of attrition is in progress.

Key Topics: **Standard topics** :
Household composition and demographic charasteristics of each individuals
Housing
Incomes (on a monthly basis)
Education
Employment, unemployment
Biography (education, employment, family background)
Life events
Determination of the Income group within the household

Special topics in single wave :
Housing background
Subjective indicators (poverty)
Difficulty to pay some expenditures
Economic behavior after a large decrease of incomes
Beneficiary of the Guaranteed Minimum Income
Project to create a self-employed activity
Non-monetary incomes
Household assets
Duration and cost of nursing
Services granted to elderly persons
Debts
Intra-family monetary transfers

References: *Présentation de l'Enquête Socio-Economique auprès des*

Ménages Lorrains (Panel Lorrain). Document d'information ADEPS, décembre 1992, 41 p., Bruno JEANDIDIER.

L'évolution de l'échantillon du panel de ménages lorrains tout au long de ses six vagues annuelles d'enquête. Les Cahiers de Recherche de l'ADEPS, n°15 , octobre 1995, 26 p., Bruno JEANDIDIER & Saïd JMEL.

Response Rates : The following table gives the raw number of persons and households for cross-sections, and of persons for 2-wave-periods, and longer periods starting in wave 2 [1].

The losses comprise losses due to field-work, non response and natural losses (death).

	no. of Households	**% with respect to start of period**	**no of individuals**	**% with respect to start of period**
cross-sections				
wave 1 : 1985	715		2609	
wave 2 : 1986	2092		7553	
wave 3 : 1987	2068		7342	
wave 4 : 1988	2100		7380	
wave 5 : 1989	2038		7115	
wave 6 : 1990	2190		7480	

[1] We choose the wave 2 because the wave 1 concerns only one third of the sample.

	no. of Households	% with respect to start of period	no of individuals	% with respect to start of period
longitudinal periods				
wave 1 - 2			2274	87.2
wave 2 - 3			6702	88.7
wave 3 - 4			6652	90.6
wave 4 - 5			6812	92.3
wave 5 - 6			6631	93.2
wave 2 - 3			6702	88.7
wave 2 - 4			6305	83.5
wave 2 - 5			5919	78.4
wave 2 - 6			5808	76.9

F.1.2 Germany

Panelname:	SOEP **So**cio-**E**conomic **P**anel
Host Institute:	The Socio-Economic Panel Deutsches Institut fuer Wirtschaftsforschung (DIW) Koenigin-Luise-Strasse 5 D - 14195 Berlin
Sponsor:	Deutsche Forschungsgemeinschaft (DFG) (German National Science Foundation)
User Support:	The Socio-Economic Panel Deutsches Institut fuer Wirtschaftsforschung (DIW) Koenigin-Luise-Strasse 5 D - 14195 Berlin Phone: + (49-30) 89 789-283 Fax:+ (49-30) 89 789-200 Internet-WWW-Page: HTTP://WWW.DIW-BERLIN.DE/SOEP/SOEPHOME.HTML Contact in the USA: Prof. Richard Burkhauser Syracuse University Maxwell School of Citizenship and Public Affairs 400 Maxwell Hall Syracuse NY 13244-1090 Phone: + (1-315) 443-9045 Fax: + (1-315) 443-1081
Cooperation:	CEPS Luxembourg / Syracuse University
Availability:	After signing a data transfer contract the SOEP-data is provided free of charge to college and university research centers. The material (diskettes or CD-rom) have to be paid for. PC and mainframe versions of the data are available.
Anonymization:	An anonymous public use file is available.
No of Waves currently available:	11

Years covered:	1984 - 1994
Subsamples:	A : West-German residents, started in 1984 B : Foreigners, started in 1984 C : East-Germans, started in 1990 D : Immigrants, not yet started
Geographic Universe:	Federal Republic of Germany (In June 1990 before the unification of Germany the survey was extended to the territory of the former GDR.)
Reference Population:	Private households Persons living in private households Institutionalized population : - not included in sample A (first wave) - included in sample B moves to institutions are followed
Case units:	Households Individuals - interviews with all household members older than 15 years - information records for children (< 16 years)
Questionnaires:	One questionnaire for each household. One questionnaire for each person. All these household and individual questionnaires also exist in the following foreign languages: greek turkish yugoslav italian spanish.
Method of Data Collection :	Face to face interview, or respondents fill in the questionnaire with or without the presence of the interviewer.
Key Topics:	Demography and Population Labor Market and Occupation Income, Taxes, and Social Security Housing Health Household Production Education, Training, and Qualification Basis Orientation (preferences and values), Participation, and Integration **and wave specific topics**

Sampling Procedure:	Sample A,C : equal probability sample of households Sample B: Disproportional sampling from foreigner registers sampling probabilty proportional to household size
Use of weights:	Estimations of population totals and related figures must be weighted in order to produce unbiased estimates. The weights comprise adjustment for the sampling procedure in wave one, and adjustments to account for attrition in later waves.
References:	Benutzerhandbuch (paper version, three volumes) User's Handbook (Word Perfect-files on CD-Rom, Syracuse) SOEP-INFO (floppy disc) menu-driven information-system SOEPLIT (floppy disc) literature based on GSOEP data Burkhauser, Richard (1991): Introduction to the German Socio-economic Panel for English Speaking Researchers Program Project Paper No. 1 Cross National Studies in Ageing, New York: Syracuse University

Response rates:
The following table shows the development of the case numbers (Total and percentages) of **households in the West-Sample of the GSOEP**. All the numbers on the diagonal are with respect to the starting-year 1984. This shows the cross-sectional view off the data-set.
All other numbers are with respect to the first number in their row and visualize the longitudinal development. Be aware that for a longitudinal analysis from year t to year t+n only households with realized interviews in all years between t and t+n are possible to use.

West - Sample	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
1984	5921 (-)	5184 (87.6)	4680 (79.0)	4425 (74.7)	4119 (69.6)	3847 (65.0)	3651 (61.7)	3492 (59.0)	3345 (56.5)	3209 (54.2)	3032 (51.2)
1985		5322 (89.9)	4798 (90.2)	4534 (85.2)	4420 (83.1)	3941 (74.1)	3740 (70.3)	3577 (67.2)	3424 (64.3)	3286 (61.7)	3106 (58.4)
1986			5090 (86.0)	4783 (94.0)	4440 (87.2)	4137 (81.3)	3924 (77.1)	3749 (73.7)	3588 (70.5)	3455 (67.9)	3248 (63.8)
1987				5026 (84.9)	4622 (92.0)	4298 (85.5)	4071 (81.0)	3888 (77.4)	3722 (74.1)	3561 (70.9)	3367 (67.0)

1988					4814 (81.3)	4466 (92.8)	4215 (87.6)	4024 (83.6)	3847 (80.0)	3675 (76.3)	3474 (72.2)
1989						4690 (79.2)	4410 (94.0)	4204 (89.6)	4011 (85.5)	3834 (81.7)	3621 (77.2)
1990							4640 (78.4)	4405 (94.9)	4198 (90.5)	4010 (86.4)	3787 (81.6)
1991								4669 (78.9)	4427 (94.8)	4218 (90.3)	3976 (85.2)
1992									4645 (78.5)	4398 (94.7)	4137 (89.1)
1993										4667 (78.8)	4364 (93.5)
1994											4600 (77.7)

Response rates:
Number of persons in the West-Sample of the GSOEP.

West - Sample	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
1984	12245 (-)	10563 (86.3)	9485 (77.5)	8864 (72.4)	8172 (66.7)	7563 (61.8)	7091 (58.0)	6725 (54.9)	6368 (52.0)	6024 (49.2)	5667 (46.3)
1985		11090 (90.6)	9941 (89.6)	9256 (83.5)	8509 (76.7)	7867 (71.0)	7369 (66.4)	6979 (62.9)	6598 (59.5)	6237 (56.2)	5872 (52.9)
1986			10646 (86.9)	9859 (92.6)	9027 (84.8)	8321 (78.2)	7779 (73.1)	7356 (69.1)	6953 (65.3)	6563 (61.6)	6173 (58.0)
1987				10516 (85.9)	9551 (90.8)	8774 (83.4)	8186 (77.8)	7739 (73.6)	7300 (69.4)	6885 (65.5)	6473 (61.6)
1988					10023 (81.9)	9190 (91.7)	8556 (85.4)	8080 (80.6)	7615 (76.0)	7175 (71.6)	6741 (67.3)
1989						9710 (79.3)	9001 (92.7)	8490 (87.4)	7986 (82.2)	7519 (77.4)	7061 (72.7)

1990							9519 (77.7)	8946 (94.0)	8402 (88.3)	7905 (83.0)	7416 (78.0)
1991								9467 (77.3)	8845 (93.4)	8307 (87.7)	7768 (82.1)
1992									9305 (76.0)	8705 (93.6)	8119 (87.3)
1993										9206 (75.2)	8540 (92.8)
1994											9001 (73.5)

Response rates:
Number of households in the East-Sample of the GSOEP.

East-Sample	1990	1991	1992	1993	1994
1990	2179 (-)	1984 (91.1)	1825 (83.8)	1699 (78.0)	1613 (74.0)
1991		2030 (93.2)	1864 (91.8)	1735 (85.5)	1645 (81.0)
1992			2020 (92.7)	1870 (92.6)	1769 (87.6)
1993				1970 (90.4)	1856 (94.2)
1994					1959 (90.0)

Response rates:
Numbers of persons in the East-Sample of the GSOEP.

East-Sample	1990	1991	1992	1993	1994
1990	4453 (-)	4033 (90.6)	3657 (82.1)	3359 (75.4)	3148 (70.7)
1991		4202 (94.4)	3804 (90.5)	3489 (83.0)	3273 (77.9)
1992			4092 (91.9)	3745 (91.5)	3502 (85.6)
1993				3973 (89.2)	3708 (93.3)
1994					3945 (88.6)

F.1.3 Hungary

Panelname: HHP (Hungarian Household Panel Study)

Host Institute: Social Research Informatics Centre (TARKI)
Sociology Department of Budapest University of Economics
Hungarian Central Statistical Office (KSH)

Sponsor: National Scientific Research Fund (OTKA)

User Support: Social Research Informatics Centre (TARKI)
Viktor Hugo u. 18-22
H-1132 Budapest
Hungary

Tel.: (36-1) 1497-531
Fax.: (36-1) 1290-470
E-mail: toth@tarki.hu
WWW: http://www.tarki.hu/index-e.html

Availability: The SPSS/PC or mainframe files are available in Hungarian and English versions from TARKI. There are no restrictions on scientific use of the data.

No of Waves currently available: 5 (available)

Years covered: 1992 to 1996

Geographic Universe: Hungary

Reference Population: Hungarian non-institutional households

Case units: Households
Individuals

Questionnaires: Household questionnaire (filled in with the help of the most competent member of the household);

Individual questionnaire - for each adult in the household

(16 years or older);

Substitute questionnaire - for each adult not available at the time of the survey (filled in with the help of the most competent member of the household).

Each questionnaire contains different blocks. Some of these blocks are wave-specific, others are not.

Method of Data Collection: Face to face interview

Sampling Procedure: A four-stage probability sample was used based on the 1990 census, stratified by county (location), settlement (size), census district (type of urbanization) and address. This nationally representative sample was created by the Central Statistical Office. The primary sampling unit was the addresses of non-institutional households. A total of 74 settlements and 437 census districts were drawn and within them a random sample of 2000 addresses were selected. An additional sample of the same size was drawn to substitute addresses falling out from the sample (unable to answer, moved away, wrong address, dead, etc.). An additional 600 households subsample covers Budapest households, making the total sample representative of Budapest also.

Use of weights: Cross-sectional weights for household and individual level are available.

Key Topics: Social Status, Wealth, Income, Economic and Financial strategies of Hungarian households; Demographic and Employment Histories (changes in labour market position) of household members

References: The First Three Waves of the Hungarian Household Panel List of Variables. TARKI, Budapest, October 1994

The Social Research Informatics Centre (TARKI) by Toth,
EURODATA
Newsletter No. 3 Spring 1996

F.1.4 **Luxembourg**

Panelname:	PSELL Panel Socio-économique / Liewen zu Lëtzebuerg
Host Institute:	CEPS / INSTEAD Centre d'Études de Population, de Pauvreté et de Politiques Socio-Économiques International Networks for Studies in Technology, Environment, Alternatives, Development
Sponsor:	State of Luxembourg
User Support:	PSELL Group at CEPS / INSTEAD B.P. 65 L- 7201 Walferdange Tel. ++352 / 33 32 33 -1
Availability:	The data are accessible in Luxembourg on a mainframe.
Anonymization:	Since the data cannot be used outside the host institute, the extent of anonymization is decided individually.
No of Waves currently available:	9
Years covered:	1985 - 1993
Time Span:	The field work campaign starts every year in May ending in October. The reference period for the interview is the twelve months subsequent to when the survey started. In this way the panel year runs from May to April, except for the first year, which took account only of the first four months of 1985.
Geographic Universe:	Grand-Duchy of Luxembourg
Reference Population:	anyone living in the Grand Duchy who is concerned by

social welfare or social protection
The basic sample represents 97% of the population living in the country.
Excluded are :
1) foreign residents who have no links with the country's social security system or who do not live in a household, where at least one of the members has such links
2) elderly persons living in a collective household such as an old people's home

Case units:

Households
Income groups within a household
Individuals

Households: A household consists of all persons who live together in a dwelling unit (house, appartment, group of rooms, or single room)
Persons within a household can be related to each other or not.

Income Groups:
In a household in which several persons have individual income, different economic arrangements are possible. An income group is a group of persons within a household, who constitute an economic unit. The concept is put into effect in accordance with strict rules.

Questionnaires:

Questionnaire A - household , one per household
Questionnaire B - groups of income , one per group of incomes located within the household

Questionnaire C and D - individual, one questionnaire per household member
C simplied individual questionnaire, for household members below the age of 16, and those who are in full-time education and have never interrupted their school education for more than one year
D all household members to whom the rules for C do not apply

Method of Data Collection:

Face to face interview

Sampling Procedure:

The initial sample is a simple random sample of persons drawn from a register from the Inspectorate General for Social Security.
In 1991 an extension was added to the sample. These households were selected in wave 1 already, but were not included in the sample by then. In 1991 these households

and their split-offs were included.

Use of weights: Estimations of population totals and related figures must be weighted in order to produce unbiased estimates. The weights comprise adjustment for the sampling procedure in wave one, and adjustments to account for attrition in later waves.
For further information, see Reference 2.

Key Topics:

household level:
dwelling, consumer durables, the household's "general" budget management and associated difficulties, measurement of poverty, for example multi-deprivation and a subjective approach to the phenomenon

income group level:
group income such as family allowances, certain forms of social welfare, the group's financial indeptedness and the difficulties encountered at financial level, wealth in the hand of group members and certain economic practices such as saving, the group's associate existence and assistance received from outside, or inversely

individual level:
age, nationality, sex, family status, short life history, schooling and/ or vocational training;
place on the labour market with a monthly timetable and a detailed description mentioning whether in active employment, on the dole or chronically unemployed plus an in-depth description of certain years;
an inventory of personal income on the basis of a monthly timetable;
medical consumption and the degree of integration into the country and adult training received;
follow-up vocational training or certain problems specific to children such as nursery facilities and school expenditure;
signposting of important events likely to change the individual's life such as changes of job, health upsets, etc.

References:

A Brief Description of PSELL / Luxembourg's Socio-Economic Panel. Research Document Nº 9102

Dispositif des pondérations des individus et des ménages de 1985 à 1992. Document PSELL Nº 63 - B. Gailly, 1994

Méthodologie générale et répertoire des variables- Année d'enquête: 1985 (Première vague). Document PSELL Nº 1 - P. Dickes, P. Hausman, A. Kerger -1987

Response Rates: The following table gives the raw number of persons and households for cross-sections, and of persons for 2-wave-periods, and longer periods starting in wave 1. The losses comprise losses due to field-work, non response and natural losses (death).

	no. of Households	% with respect to start of period	no of individuals	% with respect to start of period
cross-sections				
wave 1 : 1985	2012		6110	
wave 2 : 1986	1793		5400	
wave 3 : 1987	1644		4914	
wave 4 : 1988	1692		4868	
wave 5 : 1989	1663		4756	
wave 6 : 1990	1677		4738	
wave 7 : 1991	1957		5498	
wave 8 : 1992	1923		5363	
longitudinal periods				
wave 1 - 2			5176	84.7
wave 2 - 3			4702	87.1
wave 3 - 4			4515	91.9
wave 4 - 5			4527	93
wave 5 - 6			4371	91.9
wave 6 - 7			4443	93.8
wave 7 - 8			5102	92.8
wave 1 - 3			4507	73.8
wave 1 - 4			4160	68.1
wave 1 - 5			3896	63.8
wave 1 - 6			3587	58.7
wave 1 - 7			3400	55.6
wave 1 - 8			3200	52.4

F.1.5 **Poland**

Panelname:	PHP Polish Household Panel
Host Institute:	Department of Economics Warsaw University Dluga 44/50 St., 00-241 Warsaw Poland
Sponsor:	Central Statistical Office (CSO) Household Department
User Support:	Prof. Brunon Gorecki Department of Economics Warsaw University Dluga 44/50 St., 00-241 Warsaw Poland Tel: 004822-314725 Fax: 004822-312846
Availability:	The data are accessible at the Department of Economics (Warsaw University). The permission to access the data is given individually.
Anonymization:	The data are completely anonymized (no name and no residence code).
No of Waves currently available:	4
Years covered:	Realized: 1987 to 1990 Planned: 1993 to 1996
Geographic Universe:	Poland

Reference Population:	Persons living in private households, excluding police officers, military personnel and members of the 'nomenklatura'
Case units:	Households Individuals
Method of Data Collection:	Face to face interview
Sampling Procedure:	The data origin from a cross-sectional household budget survey of the CSO of Poland. The sampling is based on quarterly rotation of households in the yearly circle. The sampling was done once for a four year period. Annually two groups of households were surveyed. One of them (2/3 of the sample) should remain in the sample for four years, while the families in the other group (1/3 of the sample) were replaced every year by new ones. These sampling features made it possible to extract from the data sets collected for four consecutive years a subset of households surveyed through the whole four year period. The households in the subset were the candidates for a panel. However, a longitudinal use of the household surveys has never been of prime interest for the Central Statistica Office, and for this reason such a panel was never constructed and analysed in CSO. This job was done at the department of Economics, Warsaw University.
Use of weights:	Cross-sectional weights for household and individual level are available.
Key Topics:	Household composition and demographic characteristics of each individual Household incomes Individual incomes

Labour Force variables

References: Gorecki,Brunon.;Peczkowski,Marek;Andrezej Grodner: Polish Household Panel 1987-1990 as PACO Dataset , University of Warsaw 1996

F.1.6 **USA**

Panelname:	PSID Panel Study of Income Dynamics
Host Institute:	Panel Study of Income Dynamics Institute for Social Research University of Michigan Box 1248 Ann Arbor, Michigan 48106-1248 USA
Sponsor:	National Science Foundation (NSF)
User Support:	Panel Study of Income Dynamics Institute for Social Research University of Michigan Box 1248 Ann Arbor, Michigan 48106-1248 USA Phone: +(1-313) 763-5166 Fax: +(1-313) 747-4575 Internet: PSID_staff@umich.edu
Availability:	a) PSID datafiles are available to the public through Inter-University Consortium for Political and Social Research (ICPSR): Tapes and CD's b) It is possible to download the PSID data from the PSID homepage: HTTP://www.umich.edu/~psid/
Anonymization:	PSID data files are public-use files
No of Waves currently available:	25

Years covered:	1968-1994
Time Span:	PSID data collection for a given wave extends from March through September. Within the interview one set of variables is asked about the current status (current year), another set is asked about the previous calendar year, another set refer to some other particular time period specified in the set of variables. Therefore each PSID wave contains information from two years.
Subsamples:	a) SRC: a sample of households in 1968 b) SEO 2000: low-income families in 1968 c) Latino supplemental sample in 1990
Geographic Universe:	USA: Alaska and Hawaii excluded ???
Reference Population:	Households: individuals living in private Households, individuals living in institutions (prison, college dormitory, military) are excluded
Case units:	Families (Households) Individuals **Family**: A group of individuals living together who are related by blood, marriage, or adoption. **excluded:** Lodgers, conventional roommates, or employees who share the housing unit, **included:** unmarried couples, if the couple is living in a fairly permanent arrangement
Questionnaires:	Family Questionnaire Questionnaire for the head - active - unemployed/retired Questionnaire for the spouse - active - unemployed/retired Summary questionnaire for other family members
Method of Data	Telephone interviewing and in rare cases personal

Collection:	interviews, cases where telephone interviewing is problematic
Sampling Procedure:	SRC sample and SEO sample are probability samples, but the combination, however, is a sample with unequal selection probabilities
Use of weights:	Compensatory weighting is needed in estimation, at least for descriptive statistics to compensate for unequal selectionprobabilities. Weight adjustments are also needed to attempt to compensate for differentiell nonresponse
Key Topics:	**Family level:** (Only core topics) housing, family structure, geographic mobility, public assistance in form of food or housing, taxes, poverty status **Individual level:**(Only core topics) income, socioeconomic background, health, religion, military service, housework, employment information: (a) weeks worked, weeks unemployed ... (b) occupation/industry (c) work experience,(d) employment status ,(e) event history dating employment
References:	PSID User Guide PSID yearly documentation PSID newsletter Martha S. Hill: The Panel Study of Income Dynamics (Sage) Newbury London New Delhi 1992
Response Rates:	The following table gives the raw number of persons and households for cross-sections, and of persons for 2-wave-periods, and longer periods starting in wave 1.

The losses comprise losses due to field-work, non response and natural losses (death).
Since the 1983 wave was not the first wave of the PSID there are not only losses with respect to 1983.

	no. of Households	% with respect to start of period	no of individuals	% with respect to start of period
cross-sections				
wave 1 : 1983	6852	100,0	19424	100,0
wave 2 : 1984	6918	101,0	19510	100,4
wave 3 : 1985	7031	102,6	19713	101,5
wave 4 : 1986	7017	102,4	19548	100,6
wave 5 : 1987	7060	103,3	19574	100,7
longitudinal periods				
wave 1 - 2			18331	94,2
wave 2 - 3			18288	94,2
wave 3 - 4			18462	95,1
wave 4 - 5			18337	94,4
wave 1 - 3			17316	89,1
wave 1 - 4			16428	84,6
wave 1 - 5			15617	80,4

F.1.7 UK

Panel name:	BHPS British Household Panel Survey/ Living in Britain
Host Institute	ESRC Research Centre on Micro-social Change
Sponsor	Economic and Social Research Council (ESRC) and University of Essex
User Support	BHPS User Group ESRC Research Centre on Micro-social Change University of Essex Wivenhoe Park Colchester Essex CO4 3SQ Tel. +44 (0)1206 873543 Fax. +44 (0)1206 873151 e-mail. bhpsug@essex.ac.uk
Availability	ESRC Data Archive University of Essex (Free for academic use)
Confidentiality	Names, addresses, phone numbers, tax codes and day of birth are separated from substantive data in both its computer and paper records. Strict internal procedures are enforced and all staff are required to sign an undertaking of confidentiality. All external users are required to sign an Undertaking Form which specifies Conditions of Use.
Number of Waves currently available:	3
Years covered	1991-1993
Time Span	Wave One fieldwork 3rd September 1991 to 30th January 1992. Wave Two fieldwork 5th September 1992 to 30th April 1993. Wave Three fieldwork 5th September 1993 to 30th April 1994.
Geographic Universe	Great Britain (i.e. excludes N. Ireland).
Reference Population	Residents of a private households, aged 16 or over, in Great Britain. Excluding those living in institutions. An institution is defined as 'an address at which four or more unrelated people sleep; while they may or may not eat communally, the establishment must be run or managed by a person or persons

employed for this purpose by the owner'.

Case Units	Households Individuals **Households:** A household is defined as : one person living alone or a group of people who either share living accommodation OR share one meal a day and who have the address as their only or main residence.
Questionnaires	Cover sheet and enumeration grid - one per household Household questionnaire - one per household Individual questionnaire - one per adult household member (16 years or over on 31st December of survey year. Self- completion questionnaire - one per adult household member. Proxy questionnaire - a reduced version of Individual questionnaire for non-responding household members. Telephone questionnaire - similar to Proxy schedule - used from Wave 3 in exceptional circumstances. Young person's questionnaire for 11-15 year old household members - from Wave 4.
Method of Data Collection.	Face to face paper and pencil interview. CAPI planned for near future.
Sampling Procedure	The initial selection of households for inclusion in the panel survey was made using a two-stage clustered probability design and systematic sampling. The frame used for the selection of sample units was the small users Postcode Address File (PAF) for Great Britain. In the first stage of selection 250 postcode sectors were selected as the primary sampling units (PSUs) from an implicitly stratified listing of all sectors on the PAF using a systematic sampling method. In the second stage of selection, delivery points, which are approximately equivalent to addresses, were sampled from each selected PSU using an analogous systematic procedure. The sample for Wave Two and beyond consists of all eligible adults in all households where at least one interview was obtained in Wave One, regardless of whether that individual had been interviewed in Wave One. For further information see *BHPS User Manual*.
Use of weights	Weights are provided for cross-sectional and longitudinal use. Detailed notes on the derivation and use of the weights can be found in *BHPS User Manual*.

Key Topics

Household composition form : Socio-demographic characteristics of individual household members, Relationship between household members, Marital status, Household changes during past year, Geographic location

Household questionnaire: Size and Condition of dwelling Ownership status, Length of tenure, Previous ownership Household Finances Rent, Mortgage and Loan details, Local Authority and Services charges, Allowances/Rebates, Difficulties with Rent/Mortgage payments, Household Consumption, Consumer Durables, Cars, Telephones, Food Heating/Fuel types, costs, methods of payment Interview Characteristics

Individual questionnaire: Neighbourhood and Individual Demographics: Birthplace, Residence, Satisfaction with Home/Neighbourhood, Parents Employment background, Educational background and attainments, Recent Education/Training, Partisan support,
Health and Caring: Personal health condition and its effect on daily life and employment, Visits to doctor, Hospital/Clinic use, Accidents, Illness, Childbirth, NHS/Private, Use of Health and Welfare services, Social Services, Specialists, Checkups, Tests and Screening, Smoking, Attitude toward Costs/Payments for Healthcare, Childrens Health, Caring for Relatives and others in household/Outside household.
Employment: Status, Not Working/Seeking Work, Expectations,
Benefits receipt, Self Employed Contractor/Sub-contractor, Accounts, Sector- Private/Public, SIC/SOC/ISCO, Nature of Business/Duties, Workplace/Size of Firm, Travelling time/Means of travel, Length of Tenure, Hours worked/Overtime, Union membership, Prospects/Training Superannuation/Pension schemes, Attitudes to work/Incentives, Wages/Salary/Deductions, Childcare provisions - when working, costs, when children ill.

Employment History in past year: Labour Force Status, Spells, Size/Sector/Nature of Business/Duties
Wages/Salary/Deductions, Reasons for leaving/taking jobs.
Values and Opinions: Distribution of wealth and social justice, Government's Roles and responsibilities, Trade Unions, Social Class, Environment, Partisanship and interest in politics, religion, ethnicity, Membership of and activity in Social and Interest Groups.
Household finances: Incomes from Benefits, Allowances, Pensions Rents, Savings, Interest, Dividends. Management of finances now and last Year, subjective economic well-being, Pensions, Internal Transfers for Rent, Housekeeping, Board, Keep, Bills, Allowances. External Transfers for

Maintenance/Alimony/Child Support, Bills, Expenses, Education, Grants and Allowances, Loan Repayments. Personal Spending, Roles of partners/spouses in domestic work and childcare, time spent on domestic work, Financial management and bills, driving licence, car ownership and use, company car use, value of car
Self completion questionnaire: Personal Feelings: Stress, Worry, Strain, Capability, Strength, Confidence, Happiness, Unhappiness. Attitudes To Family and to Mens and Womens Roles. Social Support

Response Rates: The following table gives the raw number of individuals and households for cross-sections. For cross-wave periods only individual figures can be calculated.
The losses are due to field-work, non-response and natural losses (death).

	No. of Households	% with respect to start of period	No. of individuals	% with respect to start of period
Cross-sections				
Wave 1 1991	5511		13840	
Wave 2 1992	5227		13151	
Wave 3 1993	5528		13105	

	No. of Households	% with respect to start of period	No. of individuals	% with respect to start of period
Longitudinal periods				
Wave 1 - 2			12240	88.4
Wave 2 - 3			11978	91.1
Wave 1 - 3			11388	82.3

F. 2 Imputation methods applied to income data

PSELL / Luxembourg

In Luxembourg the following methods were applied for the imputation of income variables:

2 mean value of previous and following wave
3 inflation / deflation
4 manual method (reading the questionnaire)
5 guaranteed minimum income

The order indicates the priorities. When no change had occurred in the professional status for those persons who worked more than 10 hours / week, the mean value or inflation / deflation method was applied, depending on which information was available. Otherwise an individual or manual decision was made.
Each variable subject to imputation has an associated imputation flag variable.

BHPS / UK

Missing data on a range of income and housing cost variables have been imputed in all waves of data. Each variable subject to imputation has an associated imputation flag variable.
Two main imputation techniques were used:

Hot deck imputation:
The sample is devided into classes found to be predictive of the variable to be imputed.
Then a valid value of the variable within the same imputation class is used to impute the missing value.

Regression technique:
A regression model is fitted to all valid variables which were non-missing (or imputed) for both valid and missingcases of the variable to be imputed. The closest unit with respect to the predicted values is chosen and it's real value is used for the imputation of the missing values.

From wave two onwards cross-wave imputation was performed, which means that the value for imputation is taken from a donor who is both similar to the recipient in current characteristics and in the value of the imputed variable at the other wave.

ESEML / Lorraine:

In the Lorraine panel study missing values in income variables were imputed by using regression techniques and replacing the missing values by predicted values.
In the case of social transfers when fixed amounts were available by legislation these were taken for imputation.
Whenever the situation was unclear a manual approach (looking at the questionnaire) was made.

SOEP / Germany:

There are no imputed values in the German panel study.

PSID / USA:

Hot deck imputation was applied to income data.

G Available National Documentation

Documentation on the original panel studies and documentation on how the PACO variables were derived from the original variables are available at the CEPS institute.
This latter documentation is available in printed form and as Word-Perfect files.

Documentation on the original panel studies ::

	Written form	Meta - data on PC
PSELL	questionnaires documents on file-organization, weighting and imputation procedures	PSELLDOC : CLIPPER programme menu-driven "manual"
SOEP	handbook, incl. questionnaires	SOEPINFO : CLIPPER programme menu-driven "manual"
ESEML	questionnaires documents on file-organization	variable lists on SYBASE files
BHPS	handbook, incl. questionnaires	handbook as Word-Perfect files
PSID	handbooks	variable labels on ASCII files

In addition to the panel studies which were included in the PACO dataset, there are also panel data from Sweden and Belgium, for which documentation is also available at CEPS.